МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «Кубанский государственный
аграрный университет имени И. Т. Трубилина»

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Практикум

Краснодар
КубГАУ
2017

УДК 519.22 (076.5)
ББК 22.172
С78

Р е ц е н з е н т :

А. Г. Прудников – профессор кафедры экономического анализа Кубанского государственного аграрного университета имени И. Т. Трубилина

К о л л е к т и в а в т о р о в :

И. А. Кацко, А. М. Ляховецкий, К. Н. Горпинченко, Ю. Н. Захарова
Е. В. Кремянская, А. Е. Жминько,

С78 **Статистические методы обработки данных** : практикум / И. А. Кацко [и др.]. – Краснодар : КубГАУ, 2017. – 89 с.

Практикум содержит краткое изложение основных теоретических положений дисциплины «Статистические методы обработки данных» в тематическом разрезе, а также задания к практическим занятиям и самостоятельной работе, позволяющие сформировать и закрепить умения и навыки использования статистических методов для обработки экспериментального материала.

Предназначен для магистрантов направления «Ветеринарно-санитарная экспертиза».

УДК 519.22 (076.5)
ББК 22.172

Введение

Научные исследования в современной ветеринарии невозможны без применения современных методов статистической обработки экспериментальных данных.

Статистические методы обработки данных можно разделить на следующие группы.

1) по способу получения экспериментальных данных:

– активный эксперимент;

– пассивный эксперимент (выборочное или сплошное наблюдение).

2) по цели обработки данных:

– описательные (получение и сравнение числовых характеристик экспериментальных данных) – анализ вариационных рядов, выборочный метод, проверка статистических гипотез и другие;

– аналитические (количественная оценка и анализ зависимостей, описывающих изучаемые объекты (процессы) – дисперсионный анализ, регрессионный анализ, анализ рядов динамики и другие).

Обработка количественных данных о биологических объектах и их оценка являются задачей одного из прикладных разделов математической статистики – биологической или вариационной статистики.

Цель практикума – оказать помощь обучающимся в овладении приемами и методами статистического исследования; в закреплении теоретических знаний, полученных на лекциях и при самостоятельной работе во внеучебное время.

Настоящий практикум предназначен для магистрантов по направлению подготовки «Ветеринарно-санитарная экспертиза»».

Обучающийся, на основе изучения рекомендуемой литературы, самостоятельно выполняет задания по темам в соответствии с индивидуальным вариантом. Для облегчения выполнения самостоятельного задания, по всем темам изложены необходимые краткие методические указания и приводится решение типовых задач.

1 Вариационные ряды распределения

Результаты наблюдений − это, в общем случае, ряд чисел по отдельным единицам совокупности, расположенных в беспорядке, который для изучения и наглядного представления необходимо упорядочить. Операция, заключающаяся в расположении значений признака по возрастанию или убыванию, называется *ранжированием* данных. После операции ранжирования статистические данные можно сгруппировать так, чтобы в каждой группе признак принимал одно и то же значение, которое называется *вариантом* (x_i). Число элементов в каждой группе называется *частотой* варианта (n_i).

Признак – это характерная особенность строения или функционирования организма. Биологические признаки варьируют, но не всегда поддаются измерению.

Сумма всех частот (n) называется объемом совокупности. Отношение частоты данного варианта к объему совокупности называется *относительной частотой* (w_i) или *частостью* этого варианта.

Упорядоченная последовательность вариант изучаемого признака и соответствующих им частот (или частостей) называется *вариационным рядом* (вариация – изменение). При их составлении следует учитывать тип признаков.

Все признаки можно разделить на две группы – качественные и количественные. Типичные качественные признаки − окраска, пол, наличие заболевания и т. п. В таком случае можно без специальных измерений достаточно определенно судить о наличии или отсутствии того или иного признака у конкретной особи.

Количественные признаки можно анализировать лишь на основании специальных измерений или подсчетов. Их подразделяют на интервальные (континуальные) и дискретные. Континуальные признаки теоретически могут принимать любые возможные значения в пределах между минимальным и максимальным показателем признака. К ним относятся масса, линейные размеры и температура организма, содержание биохимических и неорганических веществ в его тканях и т. п. Дискретный – это признак, принимающий отдельные целочисленные значения, например, поголовье коров, число пятен на надкрыльях жука, яиц в гнезде.

Для построения интервального ряда весь промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений в каждом из них. Интервальный вариационный ряд может быть построен как с равными, так и не равными по длине интервалами (обычно прогрессивно возрастающими или убывающими).

Количество интервалов (групп) (k), в случае нормально распределённой, или близкой к ней, совокупности, можно определить по формуле Стерджесса:

$$k = 1 + 3{,}322\lg(n). \tag{1.1}$$

где n – численность совокупности.

Длина отдельного интервала определяется по формуле:

$$h = \frac{R}{k} = \frac{x_{max} - x_{min}}{k}. \tag{1.2}$$

где R – размах вариации.

Накопленная частота определяется путем последовательного суммирования частот вариационного ряда.

Вариационные ряды изображают графически с помощью полигона и гистограммы. В этом случае по оси абсцисс откладываются значения признака, а по оси ординат – соответствующие им частоты, частости или накопленные частоты.

Интервальный вариационный ряд изображается графически с помощью гистограммы и кумуляты распределения. На оси абсцисс откладываются границы интервалов варьирующего признака, а на оси ординат – частоты или частости. Каждому интервалу соответствует прямоугольник, по высоте равный частоте или частости.

К числовым характеристиками вариационного ряда: характеристики положения (средняя арифметическая, мода, медиана); характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации); характеристики меры скошенности (коэффициент асимметрии) и островершинности (эксцесс) распределения.

Средней арифметической ($\bar{X}$) дискретного вариационного ряда называется отношение суммы произведений вариант на соответствующие частоты к сумме частот:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum x_i n_i}{n}. \quad (1.3)$$

где x_i - значение варианты;

n_i – частота.

В интервальном вариационном ряду x_i – среднее значение i-го интервала, как полусумма его границ.

Модой (M_o) дискретного вариационного ряда называется значение признака, которое чаще всего встречается в исследуемой совокупности и имеет наибольшую частоту.

Медианой (Ме) дискретного вариационного ряда называется значение признака, которое находится в середине вариационного ряда и делит ряд пополам, при этом половина единиц совокупности имеют значения признака меньше медианного, а половина – больше.

Если дискретный вариационный ряд имеет $2n$ членов в ранжированной совокупности: $x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{2n}$, то медиана будет половиной суммы двух серединных вариант.

$$Me = \frac{x_n + x_{n+1}}{2}. \quad (1.4)$$

Если дискретный вариационный ряд в ранжированной совокупности имеет *2n+1* членов: $x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{2n}, x_{2n+1}$, то надо к сумме накопленных частот этого ряда прибавить единицу и полученное число разделить на 2.

$$Me = \frac{\sum s + 1}{2}. \quad (1.5)$$

При характеристике интервальных вариационных рядов с равными интервалами используются показатели мода и медиана:

а) медиана определяется по следующей формуле:

$$Me = x_{\mathrm{Me}} + h\frac{0{,}5n - S_{Me-1}}{n_{Me}}, \quad (1.6)$$

где x_{Me} – нижняя граница медианного интервала;

h – длина медианного интервала;

n – объем совокупности;

S_{Me-1} – накопленная частота интервала, предшествующего медианному интервалу;

n_{Me} – частота медианного интервала.

б) мода:

$$Mo = x_{Mo} + h\frac{(n_{Mo} - n_{Mo-1})}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})}, \qquad (1.7)$$

где x_{Mo} – нижняя граница модального интервала;

h – длина модального интервала;

n_{Mo} – частота модального интервала;

$n_{Mo\,-1}$ – частота интервала перед модальным интервалом;

n_{Mo+1} – частота интервала после модального интервала.

Мода и медиана используются в качестве характеристики среднего положения в случае, если границы ряда нечеткие или если ряд не симметричен.

Для изучения степени изменчивости признака используются следующие показатели вариации: размах вариации *(*$R = x_{max} - x_{min}$); дисперсия; среднее квадратическое отклонение; линейное (стандартное) отклонение; коэффициент вариации.

Размах вариации – это разность между максимальным и минимальным значением признака в совокупности. Чем больше разность между максимальным и минимальным вариантой, тем сильнее изменчивость признака.

Дисперсия дискретного ряда распределения представляет среднюю из квадратов отклонений вариант признака от средней арифметической.

$$\sigma^2 = \frac{\sum(x_i - \overline{X})^2}{n} - \textit{ простая форма,} \qquad (1.8)$$

$$\sigma^2 = \frac{\sum(x_i - \overline{X})^2 n_i}{n} - \textit{ взшенная форма}$$

Среднее квадратическое отклонение дискретного ряда распределения есть корень квадратный из дисперсии (термин введен Пирсоном в 1894 г.), представляющее собой следующее соотношение:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} - \textit{ простая форма,} \qquad (1.9)$$

$$\sigma = \sqrt{\frac{\sum(x_i-\bar{X})^2 n_i}{n}}\text{ – *взшенная форма*}$$

Очевидно, что невозможно сравнить среднее квадратическое отклонение, выраженное в кг со средним квадратическим отклонением, выраженном в процентах. С помощью σ также нельзя сравнивать изменчивость признаков в двух биологически разнородных группах. По этим причинам для сравнения изменчивости признаков выраженных в разных единицах измерения применяют коэффициент вариации, характеризующих относительную колеблемость признака совокупности.

Коэффициент вариации:

$$V = \frac{\sigma}{\bar{\bar{X}}} \cdot 100\,\%\,. \tag{1.10}$$

Понятие дисперсии являются частными случаями более общего понятия для числовых характеристик вариационных рядов – *моментов распределения.* Моментом распределения называется средняя арифметическая тех или иных степеней отклонений индивидуальных значений признака от определенной исходной величины.

Центральный момент первого порядка (нулевое свойство средней арифметической) всегда равен нулю. *Дисперсия* – это центральный момент второго порядка. Центральный момент третьего порядка служит для оценки *асимметрии распределения.* Величину асимметрии оценивают с помощью безразмерного *коэффициента асимметрии*:

$$Ka = \frac{\mu_3}{\sigma^3} = \frac{\sum(x_i - \bar{X})^3 n_i}{n\sigma^3}. \tag{1.11}$$

где μ_3 – третий момент распределения.

Коэффициент асимметрии характеризует скошенность вариационного ряда, когда частоты значений признака, равноотстоящих от средней, отличаются друг от друга. Для симметричных вариационных рядов $Ka = 0.$ При положительном значении коэффициента асимметрии преобладают варианты, больше средней арифметической, а при отрицательном значении – варианты меньшие средней арифметической, то есть обычно вариационный ряд имеет правостороннюю или левостороннюю асимметрию. Считается асимметрия существенной

Центральный момент четвертого порядка служит для оценки так называемого *эксцесса*, определяющего степень крутости (островершинности) вариационного ряда:

Эксцесс:

$$Ex = \frac{\mu_4}{\sigma^4} - 3 = \frac{\sum(x_i - \bar{X})^4 n_i}{n\sigma^4} - 3. \qquad (1.12)$$

где μ_4 – четвертый момент распределения.

Для нормального распределения эксцесс равен нулю. Положительное значение эксцесса свидетельствует об островершинности распределения по сравнению с нормальным, а отрицательное значение о плосковершинности распределения по сравнению с нормальным.

Пример 1.1. Количество яиц, отложенных волнистыми попугайчиками в зоопарке следующее: 1, 4, 5, 3, 4, 6, 2, 5, 3, 6, 5, 4, 4, 1, 3, 6, 4, 6, 6, 3, 2, 2, 2, 4. Построить дискретный вариационный ряд, определить среднее количество яиц отложенных попугайчиками в зоопарке, моду и медиану.

Решение

1. Проранжируем исходный ряд, подсчитаем частоту и частость вариант (таблица 1.1).

Таблица 1.1 – Распределение попугайчиков по числу отложенных яиц

Число яиц, x_i	Число попугайчиков, n_i	Относительная частота, w_i
1	2	2/24
2	4	4/24
3	4	4/24
4	6	6/24
5	3	3/24
6	5	5/24
$\sum$	24	1,0

2. Рассчитаем характеристики положения вариационного ряда:

$$\overline{X} = \frac{\sum x_i n_i}{n} = \frac{1 \cdot 2 + 2 \cdot 4 + 3 \cdot 4 + 4 \cdot 6 + 5 \cdot 3 + 6 \cdot 5}{24} = \frac{80}{24} = 3{,}33.$$

Учитывая, что $n_{max} = 6$, при этом $x_{mo} = 4$, следовательно $Mo = 4$;

$$Me = \frac{3+4}{2} = 3{,}5.$$

Значит, среднее количество яиц отложенных попугайчиками в зоопарке 3, наиболее часто попугайчики откладывали 4 яица, половина попугайчиков откладывали до 3-х яиц, а половина – 3 и более.

Пример 1.2. При взвешивании 60 новорожденных телят холмогорской породы получены следующие данные (кг): 27,5; 34,7; 39,3; 33,9; 36,0; 32,3; 36,7; 37,8; 33,7; 28,5; 32,3; 43,9; 32,3; 30,4; 28,3; 26,0; 25,9; 41,1; 35,3; 42,9; 28,8; 39,7; 41,5; 36,3; 30,6; 42,6; 38,4; 24,5; 32,2; 30,5; 28,8; 23,1; 39,7; 43,8; 34,9; 34,9; 34,3; 28,4; 37,3; 35,3; 36,6; 28,6; 26,0; 30,8; 37,5; 33,5; 27,7; 34,8; 35,3; 37,4; 32,1; 33,5; 32,1; 40,3; 45,1; 26,1; 40,8; 32,1; 32,0; 35,4.

Необходимо построить интервальный вариационный ряд с равными интервалами, найти относительные частоты и накопленные частоты. Рассчитать характеристики вариационного ряда.

Решение

Для определения числа групп подставим значение $n = 60$ в формулу Стерджесса: $k = 1 + 3{,}322 lg60 \approx 6{,}907; k = 7$.

Найдем длину частичного интервала:

$$h = \frac{x_{max} - x_{min}}{k} = \frac{45{,}1 - 23{,}1}{7} = \frac{22}{7} \approx 3{,}1.$$

Таблица 1.2 – Распределение новорожденных телят холмогорской породы по весу

Группы новорожденных телят холмогорской породы по весу, кг	Число телят в группе (n_i)	Накопленное число овец (S_i)	Относительная частота (w_i)
23,1–26,1	6	6	0,100
26,1–29,2	8	14	0,133
29,2–32,3	12	26	0,200
32,3–35,4	13	39	0,217
35,4–38,5	9	48	0,150
38,5–41,6	7	55	0,117
свыше 41,6	5	60	0,083
Итого	60	–	1,000

Так как длина интервала h была округлена до 3,1, то значение 45,1 оказалось вне интервала 41,6–44,7, поэтому последний интервал открытый. Получим вариационный ряд, представленный в таблице 1.2.

По формулам 1.6, 1.7 определим моду и медиану:

$$Me = 32{,}3 + 3{,}1\frac{0{,}5 \cdot 60 - 26}{13} = 33{,}25;$$
$$Mo = 32{,}3 + 3{,}1\frac{(13-12)}{(13-12)+(13-9)} = 32{,}92.$$

Значит, половина телят имеет вес до 33,25 кг, а половина – больше 33,25 кг. Наиболее часто вес телят составляет 32,92 кг.

Для расчета средней арифметической, дисперсии, коэффициента асимметрии и эксцесса построим вспомогательную таблицу 1.3.

Среднее значение признака составит:

$$\bar{X} = \frac{2009{,}2}{60} = 33{,}5 \text{ кг.}$$

Дисперсия и среднее квадратическое отклонение:

$$\sigma^2 = \frac{1719{,}55}{60} = 28{,}659,$$
$$\sigma = \sqrt{\frac{1719{,}55}{60}} = \sqrt{28{,}659} = 5{,}353.$$

Коэффициент вариации:

$$V = \frac{\sigma}{\bar{X}} \cdot 100\% = 16\ \%.$$

Таким образом, средний вес телят по исследуемой совокупности составил 33,5 кг. Вес телят в среднем колебался в промежутке $\bar{X} \pm \sigma = 33{,}5 \pm 5{,}353$, т. е. от 28,147 до 38,853 кг. Этот интервал, а так же коэффициент вариации показывают, что имеются не большие различия в весе новорожденных телят холмогорской породы.

Таблица 1.3 – Вспомогательная таблица для расчета числовых характеристик ряда распределения

Группы хозяйств по весу телят, кг	Среднее значение интервала (x_i)	Число телят группе (n_i)	$x_i n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2 n_i$	$\frac{x_i - \bar{x}}{\sigma}$	$(\frac{x_i - \bar{x}}{\sigma})^3 n_i$	$(\frac{x_i - \bar{x}}{\sigma})^4 n_i$
23,1–26,1	24,6	6	147,6	–8,9	473,84	–1,660	–27,446	8,191
26,1–29,2	27,7	8	221,6	–5,8	267,88	–1,081	–10,104	0,507
29,2–32,3	30,8	12	369,6	–2,7	86,62	–0,502	–1,517	0,000
32,3–35,4	33,9	13	440,7	0,4	2,22	0,077	0,006	2,412
35,4–38,5	37,0	9	333	3,5	111,09	0,656	2,544	20,960
38,5–41,6	40,1	7	280,7	6,6	306,15	1,235	13,197	75,866
свыше 41,6	43,2	5	216	9,7	471,74	1,814	29,866	0,427
Итого:	–	60	2009,2	–	1719,55	0,540	6,547	108,363

Коэффициент асимметрии:

$$Ka = \frac{\mu_3}{\sigma^3} = \frac{\sum (x_i - \bar{X})^3 n_i}{n\sigma^3} = \frac{6,547}{60} = 0,109.$$

Эксцесс:

$$Ex = \frac{\sum (x_i - \bar{X})^4 n_i}{n\sigma^4} - 3 = \frac{108,363}{60} - 3 = -1,194.$$

Найденное значение коэффициента асимметрии указывает, что распределение имеет небольшую правостороннюю асимметрию. Эксцесс значительно отличен от нуля, что говорит о плосковершинном распределении и возможном отличии распределения вариационного ряда от нормального распределения.

Вариационные ряды изображают графически с помощью полигона и гистограммы. В этом случае по оси абсцисс откладываются значения признака, а по оси ординат – соответствующие им частоты, частости или накопленные частоты. Полученные точки соединяются отрезками.

Полигон частот – это ломаная, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$,…, $(x_k; n_k)$.

По данным примера 1.1 построим полигон частот:

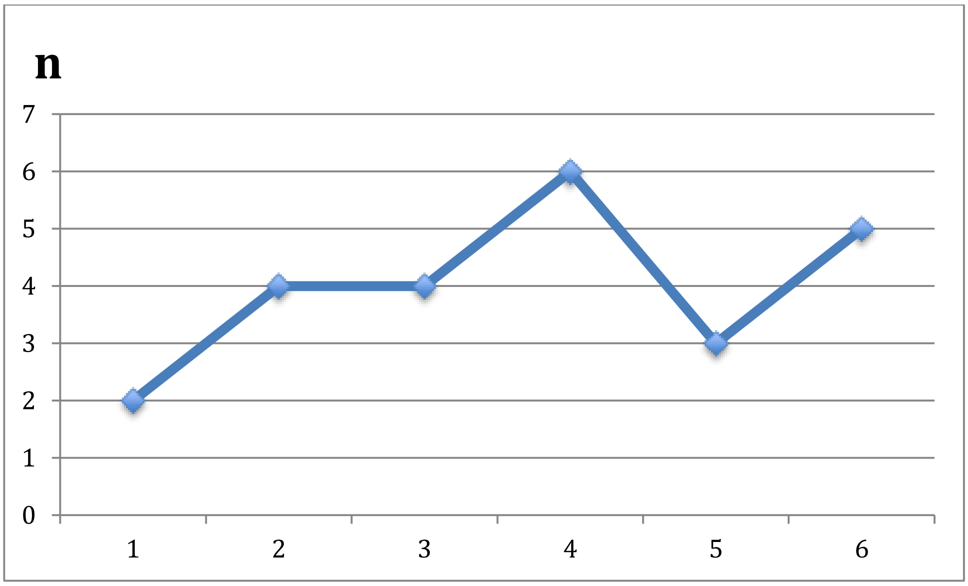

Рисунок 1.1 – Полигон распределения попугайчиков по числу отложенных яиц

Полигон относительных частот – это ломаная, отрезки которой соединяют точки: $(x_1; \frac{n_1}{n})$, $(x_2; \frac{n_2}{n})$,…, $\left(x_k; \frac{n_k}{n}\right)$. Полигон относительных частот будет иметь следующий вид (рисунок 1.2):

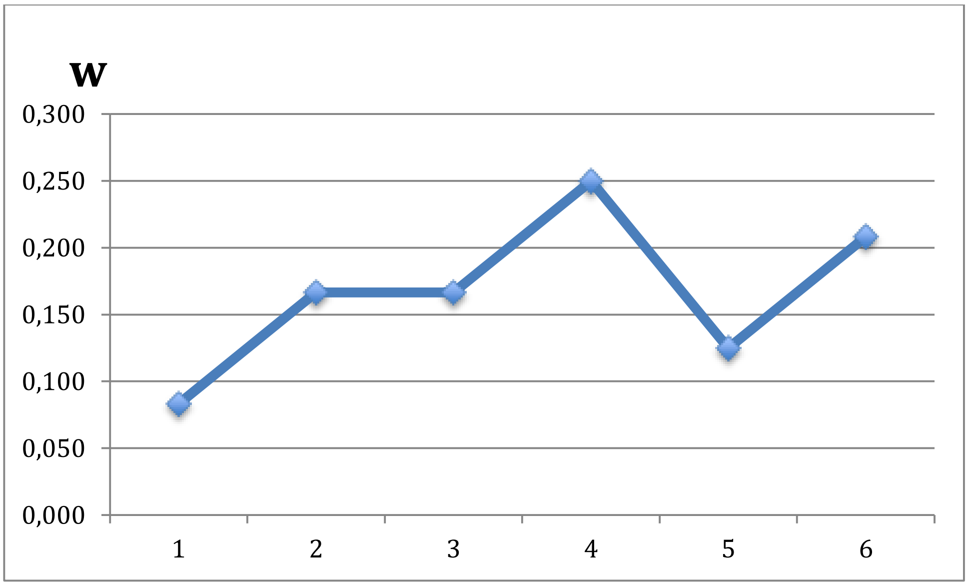

Рисунок 1.2 – Полигон относительных частот

Интервальный вариационный ряд изображается графически с помощью гистограммы и кумуляты распределения, причем открытые крайние интервалы предварительно закрываются. На оси абсцисс откладываются границы интервалов варьирующего признака, а на оси ординат – частоты или частости. Каждому интервалу соответствует прямоугольник, по высоте равный частоте или частости.

Гистограммой частот называется фигура, состоящая из прямоугольников с основанием h и высотами n_i. Для *гистограммы относительных частот* в качестве высоты рассматривают $\frac{n_i}{n}(w)$.

По данным примера 1.2 построим гистограмму частот и частостей (рисунки 1.3 и 1.4).

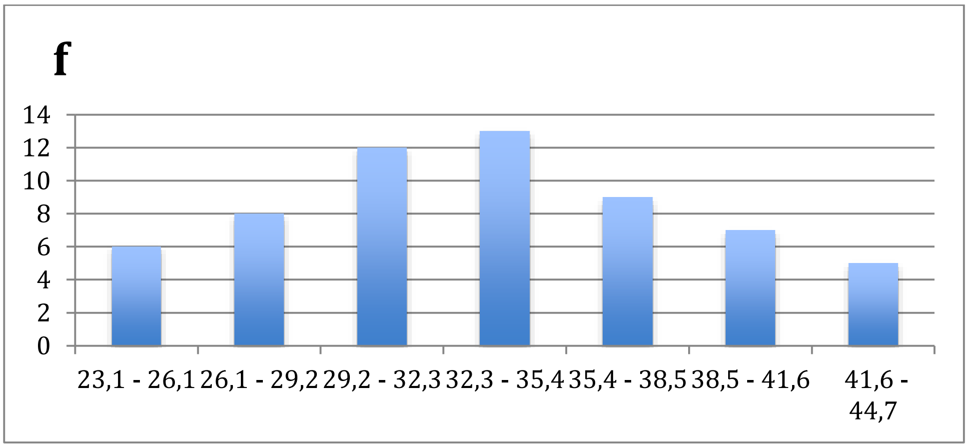

Рисунок 1.3 – Гистограмма распределения новорожденных телят холмогорской породы по весу, кг

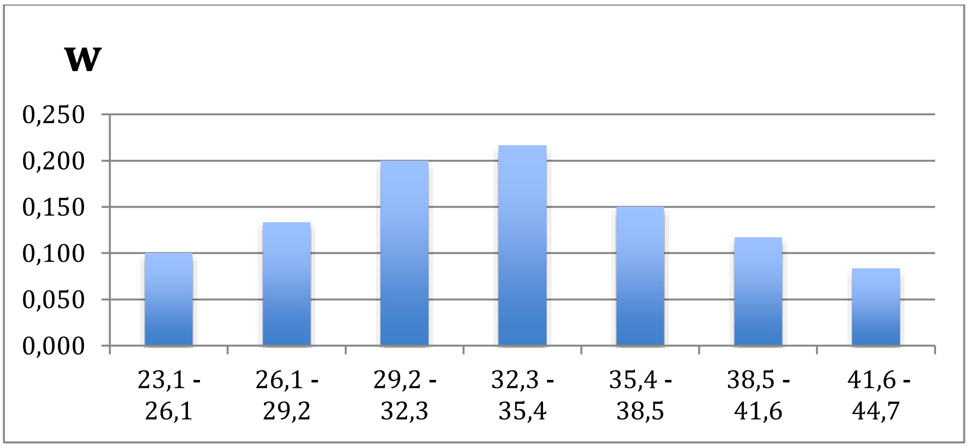

Рисунок 1.4 – Гистограмма распределения новорожденных телят холмогорской породы по весу, кг

Кумулятой называется фигура, состоящая из прямоугольников с основанием h и высотами s_i. Построим кумуляту распределения накопленных частот по данным примера 1.2 (рисунок 1.4).

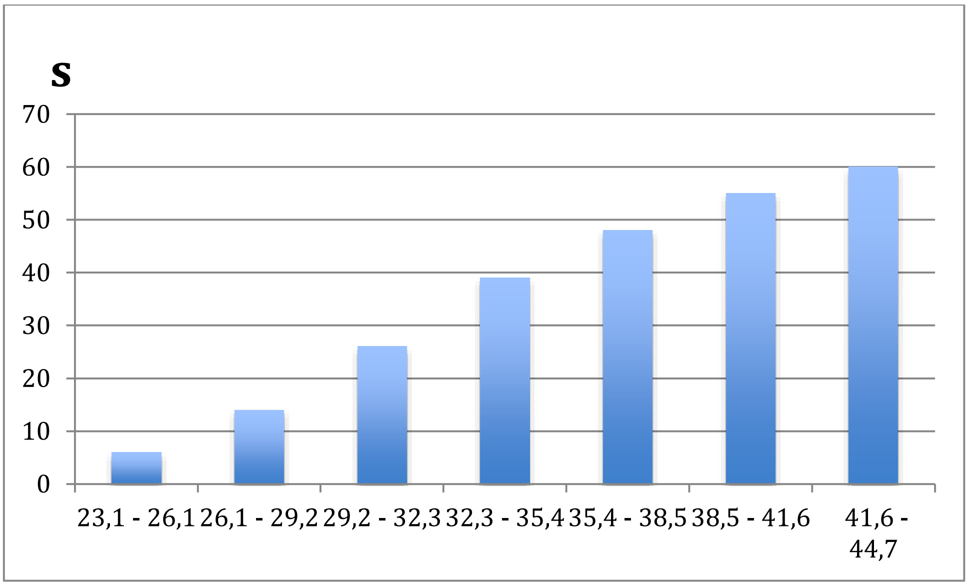

Рисунок 1.4 – Кумулята распределения новорожденных телят холмогорской породы по весу, кг

Задача 1.1. Имеется ряд распределения 80 овец Ставропольской породы по длине правого уха: 10 11 13 10 12 14 13 12 14 10 14 13 14 12 11 12 10 14 14 13 12 12 12 15 13 11 12 12 14 12 11 13 12 13 14 11 13 14 12 13 12 12 14 12 14 13 13 12 13 12 13 12 11 11 12 13 14 12 14 13 14 13 12 14 15 10 11 10 11 15 11 16 11 11 11 12 13 14 12.

Составить ряд распределения овец по длине правого уха. Найти накопленные частоты и частости. Определить средний размер правого уха, модальный и медианный размер, дисперсию и среднее квадратическое отклонение. Вариационный ряд изобразить графически.

Задача 1.2. В зоопарке за отчетный период было посчитано количество яиц, отложенных 100 волнистыми попугайчиками, которое было следующим было следующим: 4 5 4 5 5 4 5 4 3 5 6 4 4 4 5 5 3 5 5 4 6 2 3 4 5 5 5 5 5 5 6 4 6 2 5 5 3 5 5 6 4 5 5 5 5 5 5 5 4 6 7 6 3 5 5 6 5 5 4 2 4 4 6 2 6 5 4 5 5 5 4 5 4 6 5 4 7 5 6 6 4 4 4 6 5 4 3 5 5 5 5 5 4 3 7 6 4 4.

Составить ряд распределения попугайчиков по числу яиц. Найти накопленные частоты и частости. Определить среднее количество яиц, моду и медиану числа яиц, дисперсию и среднее квадратическое отклонение. Вариационный ряд изобразить графически.

Задача 1.3. В ветеринарную клинику за отчетный период с различными диагнозами поступили 110 кошек, которые имеют следующий возраст (лет): 3,5,6,4,3,4,6,4,5,3,2,2,3,4,5,3,4,5,4,1,4,5,5, 4,3,4,6,4,2,4,4,4,3,5,6,4,3,3,2,3,4,3,1,2,4,4,5,6,1,3,4,5,3,4,4,3,2,6,1,2,4,5,3,3,2,3,6,4,3,4,5,4,3,3,2,6,3,3,4,5,4,4,3,3,2,1,2,1,6,5,4,3,2,3,4,4,3,5,6,1, 5,6,4,3,4,5,6,4,3,5. Составить ряд распределения кошек по возрасту. Найти накопленные частоты и частости. Определить средний возраст кошки, модальный и медианный возраст, дисперсию и среднее квадратическое отклонение. Вариационный ряд изобразить графически.

Задача 1.4. Имеются следующие данные о диаметре фибром (доброкачественная опухоль) у крупного рогатого скота (см):

24,2; 25,8; 30,6; 32,0; 28,1; 29,6; 33,2; 34,0; 32,9; 29,9; 26,0; 28,1; 24,8; 27,3; 24,5; 30,1; 32,2; 33,6; 31,2; 31,9; 33,9; 25,6; 26,3; 27,4; 28,9; 32,0; 32,1; 33,3; 31,5; 24,9.

Составить вариационный ряд с равными интервалами, изобразить его графически. Определить средний, модальный, медианный диаметры фибром и показатели вариации.

Задача 1.5. Имеется ряд распределения числа собак породы «Такса» по числу щенят за один приплод:

Число щенят за один приплод	1	2	3	4	5
Число собак	15	220	159	67	24

Указать вид данного ряда распределения, изобразить его графически, определить среднее число щенят за один приплод, моду и медиану.

Задача 1.6. По данным таблицы 1.4 определить средний вес коз в целом и по отдельным породам. Найти накопленные частоты и частости. Вариационной ряд изобразить графически. Определить моду, медиану и показатели вариации. По одной из пород животных найти показатели асимметрии и эксцесса.

Таблица 1.4 – Численность коз в хозяйствах населения, гол.

Порода	Всего	в том числе в весе, кг				
		до 50	50–60	60–70	70–80	80 и выше
Зааненская	2517	237	639	936	664	41
Мингрельская	444	276	113	30	20	5
Оренбургская	1167	346	358	293	163	8
Тоггенбургская	2361	1282	470	347	243	19

Задача 1.7. Измерение длины предплечья у 53 особей летучей мыши вида большой трубконос Murina leucogaster Milne-Edwards без учета половой принадлежности с точностью до 0.1 мм дало следующие результаты: 38,5; 41,1; 39,6; 39,4; 39,5; 40,0; 38,6; 36,0; 39,0; 40,4; 39,5; 42,5; 41,0; 39,6; 41,8; 39,5; 40,0; 39,7; 40,0; 37,5; 42,0; 38,7; 41,5; 40,0; 42,0; 40,0; 41,5; 40,5; 37,6; 40,0; 39,5; 40,7; 40,0; 41,7; 40,0; 41,5; 41,3; 41,2; 39,0; 40,0; 40,0; 42,0; 39,3; 40,5; 39,0; 40,5; 39,5; 40,0; 39,8; 42,0; 41,2; 40,0; 39,5.

Составить статистический ряд распределения. Найти накопленные частоты и частости. Ряд распределения изобразить графически. Определить моду, медиану, среднее значение, дисперсию, среднее квадратическое отклонение, показатели асимметрии и эксцесса. Сделать выводы по результатам расчетов.

Задача 1.8. При микроскопическом анализе крови коров были получены следующие результаты (количество лейкоцитов в крови 70 коров (тыс./мм^3): 4,2 6,1 8,7 5,9 6,4 7,6 9,8 6,7 8,2 4,3 8,1 10,3 6,3 10,7 7,2 11,4 9,7 6,6 5,4 6,9 7,6 7,8 5,6 6,7 9,6 7,3 8,5 10,4 5,3 8,8 7,6 5,7 7,9 4,6 5,5 7,6 6,5 4,0 9,5 5,7 10,5 8,6 8,1 9,3 7,3 8,0 8,7 7,4 8,9 6,2 5,5 7,6 6,4 8,5 8,0 4,8 7,1 8,4 9,2 7,9 6,6 7,6 9,8 3,8 7,1 6,2 3,2 7,5 7,4 8,3.

Составить статистический ряд распределения. Найти накопленные частоты и частости. Ряд распределения изобразить графически. Определить моду, медиану, среднее значение, дисперсию, среднее квадратическое отклонение, показатели асимметрии и эксцесса. Сделать выводы по результатам расчетов.

Задача 1.9. При взвешивании новорожденных телят холмогорской породы были получены следующие результаты: 28 27 35 32 27 32 31 26 32 33 34 26 35 23 27 35 28 35 30 36 37 35 31 31 35 36 35 28 34 32 33 27 35 32 36 33 24 28 26 25 32 34 28 26 36 36 32 36 34 26 30 35 34 30 39 33 32 32 34 26 32 28 28 26 39 33 37 38 33 37 31 37 43 30 45 37 32 36 40 30.

Постройте вариационный ряд, разбив его на 7 групп. Определить моду, медиану, среднее значение, дисперсию, среднее квадратическое отклонение, показатели асимметрии и эксцесса.

Постройте вариационный ряд, разбив его на 9 групп. Определить моду, медиану, среднее значение, дисперсию, среднее квадратическое отклонение, показатели асимметрии и эксцесса.

Выясните – изменятся ли статистические показатели при изменении числа групп в вариационном ряду.

Вопросы для самоподготовки

1. Определение и виды вариационных рядов.
2. Как определяется число групп при построении вариационного ряда
3. Как определяется величина интервала при построении вариационного ряда?
4. Способы графического изображения вариационных рядов.
5. Перечислите существующие характеристики вариационного ряда.
6. Какие показатели характеризуют форму распределения?
7. Что характеризуют моменты вариационного ряда?
8. По какой методике рассчитываются коэффициенты асимметрии и эксцесс.

2 Выборочный метод

Сбор данных для статистического изучения явлений может проводиться сплошным и выборочным методами. При сплошном наблюдении обследуются все единицы изучаемой совокупности и эта совокупность называется *генеральной*. При выборочном наблюдении отбирается часть единиц генеральной совокупности, а показатели, найденные по отобранной части единиц, должны достаточно точно характеризовать показатели всей совокупности единицы.

По процедуре отбора различают два вида отбора:

– **повторный**, при котором отобранная единица возвращается назад в генеральную совокупность и может попасть в выборку более чем один раз;

– **бесповторный**, когда каждая отобранная из совокупности единица один раз участвует в процессе отбора.

При проведении выборочного наблюдения возникают ошибки регистрации и ошибки репрезентативности (представительности).

Ошибки репрезентативности – это расхождения между обобщающими характеристиками выборочной и генеральной совокупности, возникающие вследствие несплошного характера наблюдения. Желательно, чтобы величина ошибок была небольшой. Так как численное значение ошибки не известно, то ее возможная оценка дается с помощью расчета средней и предельной ошибок выборки. Обычно величина ошибок определяется для средней арифметической и для доли единиц, обладающих определенным признаком.

Средняя ошибка выборки находится в зависимости от вида и способа отбора. Различают следующие способы отбора: собственно-случайный; механический; типический (районированный); серийный (гнездовой); комбинированный; многоступенчатый; многофазный; взаимопроникающий и другие.

При простой случайной выборке отбор единиц производится из генеральной совокупности путем жеребьевки или с помощью таблицы случайных чисел. При этом способе единица наблюдения совпадает с единицей отбора.

Средняя ошибка выборки ($\mu_{\bar{x}}$) находится по формуле

а) если отбор случайный повторный:

$$\mu_x = \frac{\sigma_в}{\sqrt{n}} \qquad (2.1)$$

б) если отбор случайный бесповторный:

$$\mu_x = \sqrt{\frac{{\sigma_в}^2}{n}(1 - \frac{n}{N})} \qquad (2.2)$$

где n – объем выборочной совокупности;

N – объем генеральной совокупности;

${\sigma_в}^2$ – дисперсия выборочной совокупности.

В больших выборках (n>30) выборочная дисперсия определяется по формуле:

$$\sigma_в^2 = \frac{\sum(x_i - \tilde{x})^2 n_i}{n}, \qquad (2.3)$$

где $\tilde{x}$ – выборочная средняя.

В малых выборках ($n \leq 30$):

$$\sigma^2_{в} = \frac{\sum (x_i - \tilde{x})^2 n_i}{n-1}. \qquad (2.4)$$

Выборочная дисперсия в малых выборках обычно обозначается S^2.

Относительная средняя ошибка выборки, называемая точностью опыта находиться по формуле:

$$P\% = \frac{\mu_{\bar{x}}}{\tilde{x}}. \qquad (2.5)$$

Предельная ошибка выборки находится как предел отклонения выборочной характеристики от генеральной, гарантируемой с заданной, обычно близкой к единице, вероятностью, называемой доверительной вероятностью (P).

Предельная и средняя ошибки выборки связаны соотношением:

$$\Delta = t \cdot \mu, \qquad (2.6)$$

где μ – средняя ошибка выборки;

t – коэффициент, зависящий от уровня доверительной вероятности .

Для малых выборок t находят по таблице критических значений t-критерия Стьюдента в соответствии с уровнем доверительной вероятности и числом степеней свободы $k = n - 1$.

Обычно уровень доверительной вероятности (P) равен 0,9; 0,95 или 0,99. При большом объеме выборочной совокупности для этих уровней доверительной вероятности t равно 1,65; 1,96 или 2,58 соответственно.

Доверительный интервал, который покрывает неизвестное значение генеральной средней с заданной доверительной вероятностью, определяется неравенством:

$$\tilde{x} - \Delta_{\bar{x}} \le \bar{x} \le \tilde{x} + \Delta_{\bar{x}}, \qquad (2.7)$$

где $\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}}$ (предельная ошибка выборки).

При случайном отборе средняя ошибка выборки для доли (P) находится по формуле 2.8 и 2.9.

а) если отбор повторный:

$$\mu_p = \sqrt{\frac{w(1-w)}{n}}, \tag{2.8}$$

б) если отбор бесповторный:

$$\mu_p = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}\ . \tag{2.9}$$

В формулах (2.9 и 2.10) w – это выборочная доля единиц, обладающих данным признаком ($w = \frac{n}{N}$).

Доверительный интервал для генеральной доли определяется следующим неравенством:

$$w - \Delta_p \le P \le w + \Delta_p, \tag{2.10}$$

где $\Delta_p = t \cdot \mu_p$. (предельная ошибка доли).

Пример 2.1. Новая порода лошадей в племенном хозяйстве содержалась на 8 фермах в одинаковых условиях. Получен следующий удой молока на фермах, ц/гол.: 25,1; 28,4; 24,4; 27,6; 29,4; 27,8; 26,5; 28,8. При уровне доверительной вероятности 0,95 определить границы, в которых будет находиться средняя продуктивность лошадей в генеральной совокупности.

Решение

По условию задачи имеем: n = 8; Р = 0,95.

Найдем средний удой молока от одной кобылы и среднее квадратическое отклонение продуктивности (таблица 2.1).

$$\tilde{x} = \frac{\sum x_i}{n} = \frac{218{,}0}{8} = 27{,}25;\ \sigma^2{}_{\text{в}} = \frac{\sum(x_i - \widetilde{x})^2}{n-1} = \frac{22{,}8}{8-1} = 3{,}154;$$

$$\sigma_{\text{в}} = \sqrt{3{,}1543} = 1{,}78.$$

Отбор ферм – случайный бесповторный, но так как объем генеральной совокупности неизвестен, а отношение n / N очень мало, то расчет ошибок выборки производится по формуле для случайного повторного отбора:

$$\mu_{\bar{x}} = \frac{\sigma_{\text{в}}}{\sqrt{n}} = \frac{1{,}78}{\sqrt{8}} = 0{,}63;$$

$$P\% = \frac{\mu_{\bar{x}}}{\tilde{x}}$$

Таблица 2.1 – Вспомогательная таблица для расчета средней и дисперсии

Номер фермы	Удой молока от одной кобылы, ц/гол. (x_i)	$x_i - \tilde{x}$	$(x_i - \tilde{x})^2$
1	25,1	–2,15	4,6225
2	28,4	1,15	1,3235
3	24,4	–2,85	8,1225
4	27,6	0,35	0,1225
5	29,4	2,15	4,6225
6	27,8	0,55	0,3025
7	26,5	–0,75	0,55625
8	28,8	1,55	2,4025
Итого	218,0	–	22,08

Число степеней свободы k = n – 1 = 8 – 1 = 7. При P=0,95 и k = 7 по таблице t-критерия Стьюдента t = 2,36.

Следовательно, $\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}} = 2{,}36 \cdot 0{,}63 = 1{,}49.$

Доверительный интервал составит:

$$\tilde{x} \pm \Delta_{\bar{X}} = 27{,}25 \pm 1{,}49, \text{ то есть от } 25{,}76 \text{ до } 28{,}74.$$

Вывод. Средняя молочная продуктивность кобыл рассматриваемой породы по результатам опыта составила 27,25 ц/гол., и по фермам продуктивность в среднем колебалась в границах (27,25 ±

1,78 ц/гол.). Относительная средняя ошибка выборки, называемая точностью опыта, составила 2,3 %, что свидетельствует о достаточно высокой точности опыта. С доверительной вероятностью 0,95 можно утверждать, что средний удой молока от одной кобылы в генеральной совокупности будет находиться в интервале (27,25–1,49; 27,25+1,49), т. е. от 25,76 до 28,74 ц/гол.

Для определения **средней ошибки типической выборки** применяются формулы:

а) для средней величины количественного признака:

– при бесповторном отборе $\mu_x = \sqrt{\frac{\sigma_x^2}{n}\left(1 - \frac{n}{N}\right)}$ (2.11)

– при повторном отборе $\mu_x = \sqrt{\frac{\sigma_x^2}{n}}$ (2.12)

б) для доли признака:

– при бесповторном отборе $\mu_w = \sqrt{\frac{\sigma_w^2}{n}\left(1 - \frac{n}{N}\right)}$ (2.13)

– при повторном отборе $\mu_w = \sqrt{\frac{\sigma_w^2}{n}}$ (2.14)

Пример 2.2. Типическая пропорциональная бесповторная выборка при определении среднесуточного прироста живой массы свиней в зависимости от породы показала результаты (таблица 2.2).

Таблица 2.2 – Результаты выборки данных о среднесуточном приросте живой массы свиней

Порода свиней	Обследовано животных, гол. (n_i)	Среднесуточный прирост живой массы, г. (xi)	Дисперсия средней (σ^2),
Беркширская	325	658	21,1
Ландрас	145	529	13,2
Миргородская	215	695	23,4

Найти пределы, в которых расположен среднесуточный прирост живой массы свиней во всей совокупности, если результат следует гарантировать с вероятностью 0,954 (t = 2), а обследованию подвергалось 40 % животных.

Решение

Найдем среднесуточный прирост живой массы свиней для всех групп:

$$\overline{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{658 \cdot 325 + 529 \cdot 145 + 695 \cdot 215}{685} = 213.$$

Средняя из частных дисперсий составит:

$$\overline{\sigma^2} = \frac{\sum \sigma^2{}_i n_i}{\sum n_i} = \frac{21{,}1 \cdot 325 + 13{,}2 \cdot 145 + 23{,}4 \cdot 215}{685} = 20{,}1.$$

Средняя ошибка выборки определяется:

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{20{,}1}{685}(1 - 0{,}40)} = 0{,}13 \text{ тыс. руб.}$$

Отсюда находим предельную ошибку выборки:

$$\Delta = t\ \mu_x = 2 \cdot 0{,}13 = 0{,}26 \text{ тыс. руб.}$$

Следовательно в генеральной совокупности среднесуточный прирост живой массы свиней находится в пределах:

$$213 - 0{,}26 \leq \bar{x} \leq 213 + 0{,}26, \text{ или}$$

$$212{,}74 \leq \bar{x} \leq 213{,}26.$$

Вывод. С вероятностью 0,954 можно утверждать, что среднесуточный прирост живой массы свиней будет находится в пределах от 212,74 до 213,26 тыс. руб.

При **серийном отборе** ошибки выборки исчисляются по формулам для бесповторного отбора:

а) для средней величины количественного признака

$$\mu_x = \sqrt{\frac{\delta_x^2}{r}\left(1 - \frac{r}{R}\right)} \qquad (2.15)$$

где δ_x^2 – межсерийная дисперсия выборочной средней (δ − дельта)

$$\delta_x^2 = \frac{\sum(x_i - \bar{x})^2}{r}; \qquad (2.16)$$

б) для доли альтернативного признака

$$\mu_w = \sqrt{\frac{\delta_w^2}{r}\left(1 - \frac{r}{R}\right)}, \qquad (2.17)$$

где δ_w^2 – межсерийная дисперсия выборочной доли

$$\delta_w^2 = \frac{\sum(w_i - \bar{w})^2}{r}, \qquad (2.18)$$

где r – число серий в выборке;

R – число серий в генеральной совокупности.

Для определения ошибки выборки при повторном отборе в формулах исключается множитель $\left(1 - \frac{r}{R}\right)$.

Пример 2.3. Выборочное наблюдение цены продажи 1 упаковки гельминтала для собак более 10 кг (сироп 10 мл.) по региону проводилось при помощи отбора 5 районов. По каждому отобранному району находилась средняя цена продажи, которая оказывалась следующей (таблица 2.3).

Таблица 2.3 – Средняя цена продажи гельминтала по группе районов

Номер района	Средняя цена продажи 1 уп. гельминтала, руб.
1	240
2	250
3	245
4	255
5	262

С вероятностью 0,997 определить среднюю цену продажи по региону, где 25 районов.

Решение

Найдем общую среднюю:

$$\overline{X} = \frac{240 + 250 + 245 + 255 + 260}{5} = 250 \text{ руб.}$$

Межсерийная дисперсия составит:

$$\delta_x^2 = \frac{\sum(x_i - \bar{x})^2}{r} = \frac{(240 - 250)^2 + (250 - 250)^2 + (245 - 250)^2 + (255 - 250)^2 + (262 - 250)^2}{5} = 58{,}8$$

Средняя ошибка серийного бесповторного отбора:

$$\mu_x = \sqrt{\frac{\delta_x^2}{r}\left(1 - \frac{r}{R}\right)} = \sqrt{\frac{58{,}8}{5}\left(1 - \frac{5}{25}\right)} = 3{,}07$$

При уровне доверительной вероятности 0,997 $t = 3$.

Предельная ошибка выборки:

$$\Delta = t \cdot \mu = 3 \cdot 3{,}07 = 9{,}21 \text{ руб.}$$

Следовательно, с вероятностью 0,997 можно гарантировать, что средняя цена 1 уп. гельминтала в рассматриваемом регионе заключается в пределах:

$$250 - 9{,}21 \le \bar{x} \le 250 + 9{,}21$$
$$240{,}79 \le \bar{x} \le 259{,}79.$$

Комбинированная выборка предполагает использование нескольких способов выборки. Например, серийная (гнездовая) и случайная выборка и т. д.

При проведении выборочного наблюдения важным является обеспечение достаточно большого объема выборки, чтобы достигалась необходимая точность результатов и были приемлемы затраты средств и труда на проведение исследования.

Необходимый объем выборки (n) выводится из формул предельной ошибки выборки. При собственно-случайном повторном отборе:

$$n = \frac{t^2 \sigma_в^2}{\Delta^2}. \qquad (2.12)$$

При собственно-случайном бесповторном отборе:

$$n = \frac{t^2\sigma_в^2 N}{N\Delta^2 + t^2\sigma_в^2}. \qquad (2.13)$$

Пример 2.4. Считая полученные числовые характеристики ($\bar{x}$; σ^2) интервального ряда распределения в примере 1.2 результатом случайной бесповторной 10 % выборки, определить с доверительной вероятностью 0,95:

а) границы доверительного интервала для среднего веса телят по всей совокупности хозяйств;

б) необходимый объем выборки, если предельная ошибка будет уменьшена в 2 раза.

Решение

1. Средний вес телят по выборке n = 60 составил $\bar{x}$ =33,5 кг, дисперсия σ^2= 28,65.

Объем генеральной совокупности: $N = \frac{n}{0,1} = \frac{60}{0,1} = 600$ телят.

При доверительной вероятности 0,95 значение t = 1,96.

Тогда предельная ошибка выборки составит:

$$\Delta_{\bar{x}} = 1,96\sqrt{\frac{28,65}{60}\left(1 - \frac{60}{600}\right)} = 1,285.$$

$$\tilde{x} \pm \Delta_{\bar{X}} = 33,5 \pm 1,3, \text{ то есть от 32,2 до 34,8.}$$

2. Необходимый объем выборки при предельной ошибке, уменьшенной в два раза, будет равен:

$$n = \frac{t^2\sigma^2 N}{N\Delta^2 + t^2\sigma^2} = \frac{1,96^2 \cdot 28,65 \cdot 600}{600 \cdot 0,643^2 + 1,96^2 \cdot 28,65} = 180.$$

Вывод. Средний вес телят во всей генеральной совокупности при доверительной вероятности 0,95 определяется промежутком 33,5 ± 1,3 кг, т. е. покрывается интервалом от 32,2 до 34,8 кг. Необходимый объем выборки n = 180 телят, т. е. при уменьшении предельной ошибки в 2 раза, объем выборки увеличивается в 3 раза (180 : 60 = 3).

Задача 2.1. В ветеринарной клинике в отчетном году находилось на лечении 460 собак. В результате случайного бесповторного отбора обследовано 90 животных, у 30 % из которых выявлена болезнь Карре. При уровне доверительной вероятности 0,997 определить долю и количество собак, заболевших чумкой, в целом по клинике.

Задача 2.2. Типическая пропорциональная бесповторная выборка при определении среднесуточного прироста живой массы свиней в зависимости от породы показала результаты (таблица 2.4).

Таблица 2.4 – Результаты выборки данных о среднесуточном приросте живой массы свиней

Порода крупного рогатого скота	Обследовано животных, гол. (n_i)	Среднесуточный прирост живой массы, г (xi)	Дисперсия средней (σ^2)
Симментальская	542	787	18,9
Черно-пестрая	231	659	23,6
Голштинская	428	712	21,7

Найти пределы, в которых расположен среднесуточный прирост живой массы свиней во всей совокупности, если результат следует гарантировать с вероятностью 0,954, а обследованию подвергалось 35 % животных.

Задача 2.3. Выборочное наблюдение цены продажи 1 упаковки вакдерм-F по региону проводилось при помощи отбора 5 районов. По каждому отобранному району находилась средняя цена продажи, которая оказывалась следующей: 1-й район – 118 руб./уп.; 2-й район – 135 руб./уп.; 3-й район – 144 руб./уп.; 4-й район – 123 руб./уп. ; 5-й район – 137 руб./уп.

С вероятностью 0,997 определить среднюю цену продажи по региону, где 18 районов.

Задача 2.4. Проводилось испытание 10 пород коров в крупном агрохолдинге региона (таблица 2.5). Каждая порода выращивалась на 6 фермах в равных условиях. По одной породе определить среднюю продуктивность коровы, среднюю и предельную ошибку выборки. Уровень доверительной вероятности принять 0,95.

Таблица 2.5 – Удой молока от одной коровы, ц/гол

Номер фермы	Номер породы									
	1	2	3	4	5	6	7	8	9	10
1	55,1	49,4	60,2	41,2	55,6	66,5	78,1	49,6	66,4	54,2
2	50,6	51,1	61,3	44,1	54,8	68,8	78,2	55,4	67,2	60,3
3	54,2	45,7	59,8	49,6	49,7	70,2	70,3	53,2	75,4	62,1
4	58,7	47,9	64,1	43,5	52,3	64,7	79,2	56,7	73,1	59,6
5	53,4	50,4	65,4	46,7	54,2	68,4	80,6	54,2	66,2	58,6
6	56,8	51,2	63,8	45,2	50,7	67,1	81,2	50,8	69,1	61,4

Задача 2.5. В результате случайного бесповторного отбора 200 овец из 2130 голов определено, что средний возраст составил 38 месяцев при среднем квадратическом отклонении 10 месяцев, доля ярок равна 0,35. С вероятностью 0,954 определить границы среднего возраста овец в стаде и удельного веса ярок.

Задача 2.6. В районе имеется 120 крестьянских (фермерских) хозяйств. Сколько хозяйств необходимо взять для обследования, если известно, что среднее поголовье свиней составляет 103 гол., при среднем квадратическом отклонении 18 гол. Уровень вероятности принять 0,95, точность 5 %.

Задача 2.7. Для определения влияния микроэлементов на результаты откорма свиней проведен опыт по 8 рационам животных. Рационы отличаются набором и дозами микроэлементов (таблица 2.4).

Таблица 2.4 – Результаты откорма свиней в опыте

Рацион	Поголовье свиней, гол.	Среднесуточный прирост живой массы, г	Среднее квадратическое отклонение, г
1	90	500	40
2	75	575	45
3	100	610	54
4	50	450	52
5	70	590	65
6	60	650	70
7	110	490	48
8	80	540	62

С доверительной вероятностью 0,95 определить границы, в которых будет находиться среднесуточный прирост свиней по каждому рациону и по опыту в целом.

Задача 2.8. Средний процент жира в молоке за лактацию коров холмогорских помесей был следующим: 3,4; 3,6; 3,2; 3,1; 2,9; 3,7; 3,2; 3,6; 4,0; 3,4; 4,1; 3,8; 3,4; 4,0; 3,3; 3,7; 3,5; 3,6; 3,4; 3,8. С доверительной вероятностью 0,95 и 0,99 определить границы, в которых будет находиться средний процент жира.

Задача 2.9. Определить необходимую численность выборки поголовья коров (*n*) для выборочного обследования среднегодового удоя молока от 1 коровы при повторном способе отбора, если предельная ошибка выборки (Δ_x) не должна превышать 1 ц при доверительной вероятности $P(t)$ = 0,950. Колеблемость продуктивности σ = 8 ц.

Вопросы для самоподготовки

1. Какое наблюдение называется выборочным?
2. В чем преимущества выборочного метода в сравнении с другими видами статистического наблюдения?
3. Какие вопросы необходимо решить для проведения выборочного наблюдения?
4. Что означает ошибка репрезентативности, какие факторы определяют ее величину?
5. Каковы условия правильного отбора единиц совокупности при выборочном наблюдении?
6. Как производятся собственно-случайный, механический, типический и серийный отборы?
7. В чем различие повторной и бесповторной выборки?
8. Что определяет собой средняя ошибка выборки (для средней и доли)?
9. По каким расчетным формулам находят средние ошибки выборки (для средней и доли) при повторном и бесповторном отборе?
10. Что характеризует предельная ошибка выборки и по каким формулам она исчисляется?
11. Что показывает коэффициент доверия?
12. По каким формулам определяется необходимая числен-

ность выборки, обеспечивающая с определенной вероятностью заданную точность наблюдения?

3 Корреляционно-регрессионный анализ

В большинстве исследований животных, требуется выяснить связь между различными признаками, например, для животноводства может представлять интерес связь между жирностью молока и его белковостью у коров определенной породы, связь уровня удоя с уровнем кормления, связь между кровностью помесей и их продуктивностью и т. п.

В природе могут быть обнаружены связи двух типов: функциональные и корреляционные. Функциональные связи чаще обнаруживаются в физических и химических процессах. В биологии, и следовательно, в животноводстве, специалисты чаще встречаются с корреляционными зависимостями.

Функциональной называется связь между переменными, когда каждому значению одной переменной величины, соответствует вполне определенное значение другой переменной и наоборот, т. е. $y = f(x)$.

При корреляционной связи с изменением факторного признака (признаков) на определенную величину изменяется среднее значение результативного признака. Факторными (объясняющими, независимыми) признаками называются признаки или переменные, оказывающие влияние на другие признаки. Они могут быть случайными и неслучайными.

Результативными (объясняемыми, зависимыми) называются признаки, формирующиеся под влиянием факторных признаков.

Прямолинейная связь между переменными выражается уравнением прямой на плоскости, или в пространстве или в гиперпространстве. *Криволинейные* связи выражаются уравнениями кривых различного вида: парабола, гипербола, показательная, степенная, логистическая и прочие (с увеличением возраста коров удой за лактацию в начале повышается до 5–8 отела, а затем по мере дальнейшего увеличения возраста он начинает снижаться).

Корреляционные связи могут выявлять связь не только между *количественными* признаками (удой и питательность рациона), но и между *качественными* признаками (тип рациона родителей и пол потомства).

Простая корреляционная связь – это связь меду двумя признаками, без учета других существующих связей (например, связь кормления и удоя)

$$Y = f(X).$$

Множественная корреляционная связь выясняет связь сразу между несколькими показателями (удой, порода, возраст, живой вес)

$$Y = f(X_1,\ X_2, \dots, X_n).$$

При построении уравнения множественной регрессии обычно используются следующие функции:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots .. + b_px_p\ ,$$
$$y = b_0x_1^{b_1}x_2^{b_2} \dots x_p^{b_p},$$
$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_2^2 + b_5x_1x_2.$$

С помощью корреляционного анализа оценивается *направление* и *теснота* связи между изучаемыми переменными. Его задачами являются:

1) выбор наиболее приемлемого показателя тесноты связи между переменными (коэффициент корреляции, корреляционное отношение, ранговый коэффициент корреляции, коэффициент конкордации, коэффициент взаимной сопряженности и т. п.);

2) точечная и интервальная оценка показателя тесноты связи по выборочным данным;

3) статистическая проверка значимости показателя тесноты связи;

4) формулирование вывода о наличии или отсутствии связи между переменными.

С корреляционным анализом тесно связан регрессионный анализ. Их объединяют методы обработки данных, отличают цели и формы установления связи.

Регрессионный анализ заключается в выборе и обосновании математического уравнения (совокупности уравнений), выражающего аналитически зависимость между признаками. К основным задачам регрессионного анализа относятся:

1) определение аналитического вида функции, описывающей связь между результативным и факторными признаками;

2) нахождение параметров уравнения связи;

3) определение теоретических значений результативного признака по каждой единице совокупности при фактических значениях факторных признаков;

4) нахождение отклонений фактически наблюдаемых значений результативного признака от теоретических значений;

5) оценка значимости параметров и всего уравнения регрессии.

Коэффициент корреляции определяется по выборочным наблюдениям, поэтому возникает задача оценки его статистической значимости. Рассматривается нулевая гипотеза – коэффициент корреляции равен нулю в генеральной совокупности, т.е. не является статистически значимым ($H_0 : r = 0$) и альтернативная – коэффициент корреляции существенно отличен от нуля ($H_1 : r \neq 0$) в генеральной совокупности. Проверка гипотезы чаще всего осуществляется по t-критерию Стьюдента. Критическое значение $t_{кр}$ находится по таблице Приложения Г для двусторонней области при заданном уровне значимости α и числе степеней свободы $k = n - 2$. Наблюдаемое значения критерия определяется по формуле:

$$t_{\text{набл}} = |r|\sqrt{\frac{n-2}{1-r^2}}. \qquad (3.1)$$

Если $t_{\text{набл}} > t_{\text{кр}}$, то нулевая гипотеза отвергается, коэффициент коррелляции считается статистически значимым в генеральной совокупности. Если же $t_{\text{набл}} < t_{\text{кр}}$, то нулевая гипотеза принимается, коэффициент корреляции может быть равен нулю в генеральной совокупности.

Коэффициент корреляции характеризует тесноту связи между результативным признаком и одним фактором. Но на результативный признак могут оказывать влияние несколько факторов. В этом случае целесообразно изучать *множественную корреляцию*, используя *множественные* или *частные коэффициенты корреляции.*

Корреляционным отношением Y на X называется отношение межгруппового среднего квадратического отклонения δ_y переменой Y к её общему среднему квадратическому отклонению σ_y .

$$\eta_{yx} = \frac{\delta_y}{\sigma_y}, \tag{3.2}$$

где η_{yx} - корреляционное отношение Y на X ($\eta -$ эта).
Межгрупповая дисперсия определяется по формуле:

$$\delta_y = \frac{\sum_i (\bar{Y}_i - \bar{Y})^2}{n}. \tag{3.3}$$

Аналогично определяется корреляционное отношение X на Y. Основные свойства корреляционных отношений:

1) $0 \le \eta_{yx} \le 1$;

2) если $\eta = 0$, то корреляционная связь отсутствует;

3) если $\eta = 1$, то переменные связаны функционально;

4) для линейной зависимости между переменными X и Y необходимо и достаточно, чтобы выполнялось равенство $|r| = \eta_{yx}$;

5) $\eta_{xy} \neq \eta_{yx}$;

6) $|r| \le \eta$.

Пример 3.1. Имеются следующие выборочные данные по 15 хозяйствам региона (таблица 3.1). По данным об удое молока от одной коровы и расходу концентрированных кормов на одну голову требуется:

1. Построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии, используя следующие функции: а) линейную; б) степенную; в) экспоненциальную; г) показательную.

2. Рассчитать параметры уравнения регрессии различными методами.

3. Оценить качество каждого уравнения с помощью средней ошибки аппроксимации.

4. Найти коэффициент эластичности.

5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.

6. Оценить, для линейной функции, значимость коэффициентов корреляции и регрессии по t-критерию Стьюдента при уровне значимости $\alpha = 0,05$.

7. Охарактеризовать статистическую надежность результатов регрессионного анализа с использованием F-критерия Фишера при уровне значимости $\alpha = 0,05$.

8. Определить прогнозное значение результативного признака, для линейной функции, если возможное значение факторного признака составит 1,1 от его среднего уровня по совокупности.

Таблица 3.1 – Удой молока от одной коровы и расход концентрированных кормов на одну голову

№ п/п	Расход кормов на 1 гол., ц корм. ед. (x)	Продуктивность, ц/гол. (y)
1	38,4	62,3
2	24,2	30,1
3	29,2	47,3
4	23,0	29,9
5	18,2	37,2
6	33,2	46,1
7	14,1	22,3
8	26,2	43,0
9	20,1	34,1
10	35,0	49,2
11	31,7	41,4
12	24,4	37,4
13	18,9	28,2
14	27,1	37,0
15	17,0	26,1

Решение

1. Регрессия в виде линейной функции имеет вид:

$$\hat{y} = a + bx$$

Продуктивность коров зависит от расхода концентрированных кормов, поэтому факторным признаком является количество расхода кормов на 1 гол., а результативным признаком – удой молока на одну голову (рисунок 3.1).

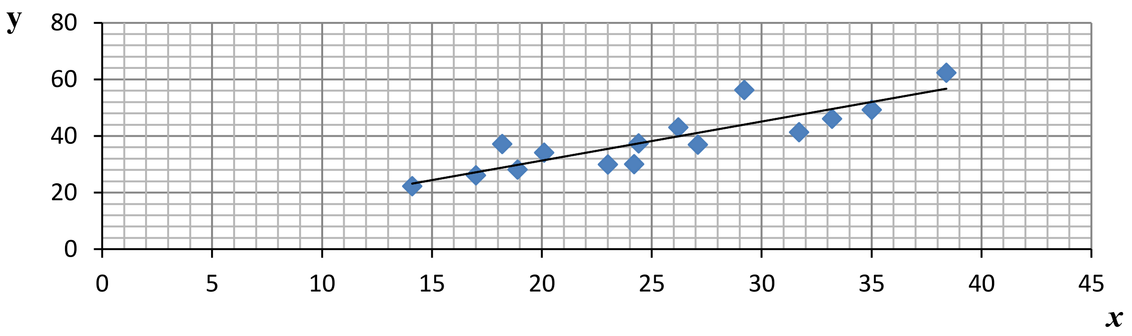

Рисунок 3.1 – Зависимость продуктивности коров (ц/гол.) от расхода кормов на 1 голову (ц. корм. ед.)

Построим график зависимости переменных х и у в прямоугольной системе координат. Характер расположения точек на графике показывает, что связь между переменными может выражаться линейным уравнением регрессии.

Для проведения всех расчетов построим вспомогательную таблицу 3.2.

В таблице все средние находятся по формуле средней арифметической простой: $\bar{x} = \Sigma x : n$.

Замечание. Расчет вспомогательной таблицы можно осуществить в табличном процессоре Excel. Результаты вычисления округлены.

1. Найдем параметры уравнения регрессии:

а) методом наименьших квадратов.

$$\begin{cases} \Sigma y = na + b\Sigma x, \\ \Sigma yx = a\Sigma x + b\Sigma x^2. \end{cases} \qquad (3.4)$$

Подставим полученные суммы в систему уравнений, учитывая, что n = 15.

$$\begin{cases} 571,6 = 15a + 380,7b, \\ 15449,92 = 380,7a + 10370,45b. \end{cases}$$

получим: a = 4,326; b= 1,331.

Таблица 3.2 – Вспомогательная таблица регрессионного анализа для уравнения линейной регрессии

№ п/п	x	y	x^2	y^2	xy	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$	$A=\left\|\frac{y-\hat{y}}{y}\right\|\cdot 100\%$
1	38,4	62,3	1474,56	3881,29	2392,32	55,436	6,864	47,109	11,017
2	24,2	30,1	585,64	906,01	728,42	36,536	–6,436	41,425	21,383
3	29,2	47,3	852,64	2237,29	1381,16	43,191	4,109	16,882	8,687
4	23,0	29,9	529,00	894,01	687,70	34,939	–5,039	25,392	16,853
5	18,2	37,2	331,24	1383,84	677,04	28,550	8,650	74,819	23,252
6	33,2	46,1	1102,24	2125,21	1530,52	48,515	–2,415	5,833	5,239
7	14,1	22,3	198,81	497,29	314,43	23,093	–0,793	0,629	3,557
8	26,2	43,0	686,44	1849,00	1126,60	39,198	3,802	14,454	8,841
9	20,1	34,1	404,01	1162,81	685,41	31,079	3,021	9,126	8,859
10	35,0	49,2	1225,00	2420,64	1722,00	50,911	–1,711	2,928	3,478
11	31,7	41,4	1004,89	1713,96	1312,38	46,519	–5,119	26,201	12,364
12	24,4	37,4	595,36	1398,76	912,56	36,802	0,598	0,357	1,598
13	18,9	28,2	357,21	795,24	532,98	29,482	–1,282	1,643	4,546
14	27,1	37,0	734,41	1369,00	1002,70	40,396	–3,396	11,533	9,179
15	17,0	26,1	289,00	681,21	443,70	26,953	–0,853	0,728	3,268
Итого	380,7	571,6	10370,45	23315,56	15449,92	–	–	279,058	142,119
Среднее значение	25,38	38,107	691,36	1554,37	1029,99	–	–	18,604	9,475

б) параметры уравнения регрессии также можно найти по формулам, вытекающим из системы нормальных уравнений:

$$b=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\left(\overline{x}\right)^2},$$

$$a=\overline{y}-b\cdot\overline{x}. \qquad (3.5)$$

Небольшие расхождения в результатах расчетов могут происходить за счет округления средних значений во втором случае.

Таким образом, уравнение регрессии имеет вид:

$$\hat{y}=4{,}326+1{,}331x.$$

Коэффициент регрессии показывает, при увеличении количества расхода концентрированного кормов на одну голову на 1 ц корм. ед., удой молока увеличивается в среднем на 1,331 ц/гол.

Если в уравнение регрессии подставить фактические значения переменной х, то определяются возможные (теоретические) значения переменной ŷ, которые наносятся на график в виде прямой.

3. Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации

$$\overline{A} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| .100\ \%;$$

$$A = \frac{142{,}199}{15} = 9{,}475. \tag{3.6}$$

Значит, фактические значения удоя молока от одной коровы различают от расчетных по уравнению регрессии в среднем различаются на 9,475 %.

Качество уравнения регрессии считается хорошим, если ошибка аппроксимации не превышает 8–10 %. Полученное уравнение регрессии можно оценить как вполне хорошее.

4. При линейной форме связи, средний коэффициент эластичности находится по формуле:

$$Э = b \cdot \frac{\overline{x}}{\overline{y}}, \tag{3.7}$$

где $\overline{x}$ и $\overline{y}$ – средние значения признаков.

$$\bar{Э} = 1{,}331 \cdot \frac{25{,}380}{38{,}107} = 0{,}886$$

Коэффициент эластичности показывает, что при увеличении количества расхода концентрированного кормов на одну голову на 1 %, удой молока от одной коровы в среднем возрастает на 0,886%.

5. При линейной зависимости теснота связи между переменными х и у определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}, \tag{3.8}$$

где σ_x и σ_y – средние квадратические отклонения по x и y.

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{691{,}363 - 25{,}38^2} \approx 6{,}872;$$
$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{1554{,}371 - 38{,}107^2} \approx 10{,}112;$$
$$r = \frac{1029{,}995 - 25{,}380 \cdot 38{,}107}{6{,}872 \cdot 10{,}112} \approx 0{,}904.$$

По величине коэффициента корреляции судят о тесноте связи. Имеется следующая градация:

$0 \le r \le 0{,}2$ - очень слабая;
$0{,}2 \le r \le 0{,}5$ - связь слабая;
$0{,}5 \le r \le 0{,}75$ - связь средняя;
$0{,}75 \le r \le 0{,}95$ - связь сильная, тесная;
$0{,}95 \le r \le 1$ - функциональная.

Так как значение коэффициента корреляции близко к единице, то между признаками связь довольно тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0{,}904^2 = 0{,}817$ показывает, что 81,7 % различий в продуктивности коров объясняется вариацией расхода кормов, а 18,3 % другими, неучтенными факторами.

6. Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции. Выдвигаем нулевую гипотезу: коэффициент корреляции в генеральной совокупности равен нулю и изучаемый фактор не оказывает существенного влияния на результативный признак: $H_0 : r_г = 0$, при $H_1 : r_г \neq 0$.

Для проверки нулевой гипотезы применим t-критерий Стьюдента.

Найдем наблюдаемое значение t-критерия

$$t_н = |r|\sqrt{\frac{n-2}{1-r^2}} = 0{,}904\sqrt{\frac{15-2}{1-0{,}904^2}} \approx 7{,}62.$$

Критическое значение t находится по таблицам распределения t-критерия Стьюдента при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы к = n – 2 = 15 – 2 = 13 для двухсторонней критической области, $t_{кр}$ = 2,16.

Сравниваем $t_н$ с $t_{кр}$. Так как $t_н > t_{кр}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, расход кормов на 1 гол. оказывает статистически существенное влияние на продуктивность коров.

Статистическая значимость коэффициента регрессии также проводится с использованием критерия t-критерия Стьюдента.

Находится наблюдаемое значение критерия:

$$t_н = \frac{b}{m_b};\ m_b = \sqrt{\frac{\sum(y-\hat{y})^2}{(n-2)\times\sum(x-\bar{x})^2}} = \sqrt{\frac{\sum(y-\hat{y})^2}{(n-2)\times\sigma_x^2\times n}} = \sqrt{\frac{286{,}374}{(15-2)\times 6{,}872^2\times 15}} \approx 0{,}176;$$

$$t_н = \frac{1{,}331}{0{,}176} \approx 7{,}56.$$

Критическое значение t также равно 2,16. Так как $t_н > t_{кр}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния расхода кормов на продуктивность коров.

7. Статистическая надежность уравнения регрессии проверяется с использованием критерия F-критерия Фишера – рассматривается нулевая гипотеза $H_0 : r^2 = 0$, при альтернативной $H_0 : r^2 \neq 0$ (или нулевая гипотеза $H_0 : b = 0$, при $H_1 : b \neq 0$). Наблюдаемое (фактическое) значение F-критерия находится по формуле:

$$F_н = \frac{\sum(\hat{y}-\bar{y})^2/m}{\sum(y-\hat{y})^2/(n-m-1)}; \qquad (3.9)$$

где m – число параметров при переменных x;

n – число наблюдений.

Если применяется линейное уравнение регрессии, то расчет $F_н$ упрощается.

$$F_н = \frac{r^2}{1-r^2}\cdot(n-2) = \frac{0{,}817}{1-0{,}817}\cdot 13 = 58{,}03$$

При уровне значимости α = 0,05 и числе степеней свободы $k_1 = m = 1,\ k_2 = n - m - 1 = 15 - 1 - 1 = 15 - 2 = 13$ по таблице находится критическое значение F-критерия.

$$F_{кр} = F_{\alpha=0,05}(k_1 = 1, k_2 = 13) = 4,67.$$

Так как $F_н > F_{кр}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты, так как $t_b^2 = t_r^2 = F$ (наблюдаемые отличия объясняются ошибками округлений).

8. Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (xp).

По условию: $x_p = \bar{x} \cdot 1,1 = 25,380 \cdot 1,1 = 27,918$.

Тогда прогнозное значение удоя молока от одной коровы составит:

$$\hat{y}_p = a + bx_p = 4,326 + 1,331 \times 27,918 = 41,485.$$

Значит, при расходе кормов на 1 гол. в 27,918 ц корм ед. возможный удой молока от одной коровы составит 41,485 ц/гол.

Замечание. Важнейшим методом анализа данных является визуализация (представление данных в виде таблиц, диаграмм, кросс-таблиц, кросс-диаграмм, графиков). Рассмотрим применение диаграммы рассеяния. Выделим в Excel диапазон B2:C16, выполним команду: Вставка – Точечная – Точечная с маркерами. В результате получим рисунок 3.2.

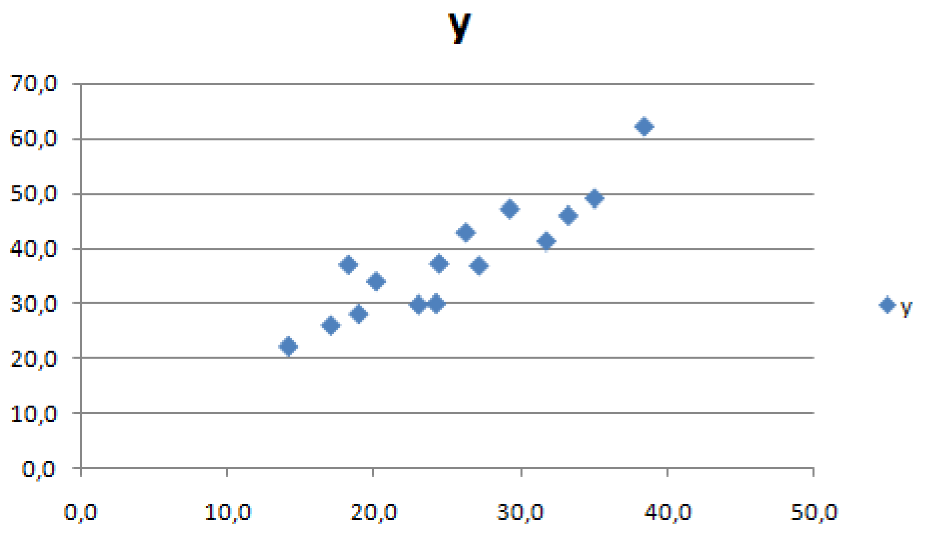

Рисунок 3.2 – Диаграмма рассеяния

Важность графического представления данных заключается в возможности увидеть возможные ошибки, допущенные при вводе данных (артефакты – объекты созданные человеком) или неоднородные значения признаков – выбросы – явно не принадлежащие изучаемой совокупности.

Например, при вводе исходных данных мы вместо 62,3 ввели 623. Построим соответствующую диаграмму рассеяния (рисунок 3.3) из которой видно, что есть наблюдение, отличающееся от других данных.

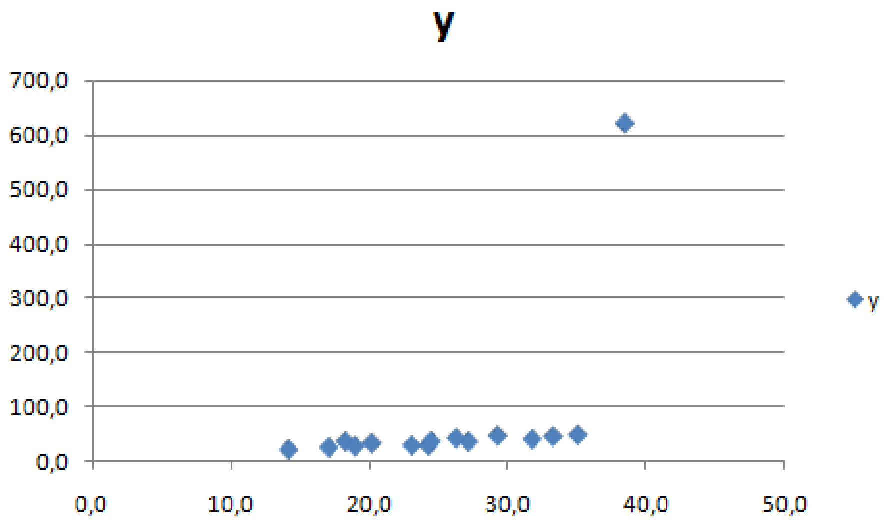

Рисунок 3.3 – Диаграмма рассеяния с артефактом (или выбросом)

Важным методом анализа данных в Excel являются диаграммы.

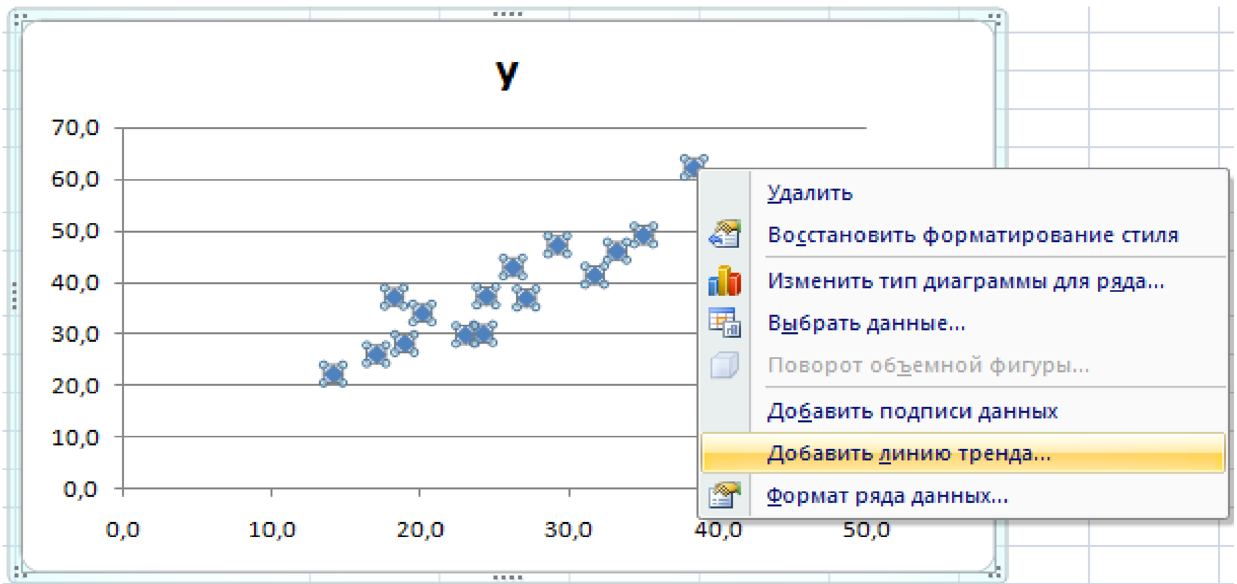

Рисунок 3.4 – Контекстное меню выделенных точек наблюдений

Выделим на рисунке 3.4 щелчком левой клавиши мыши маркеры наблюдений.

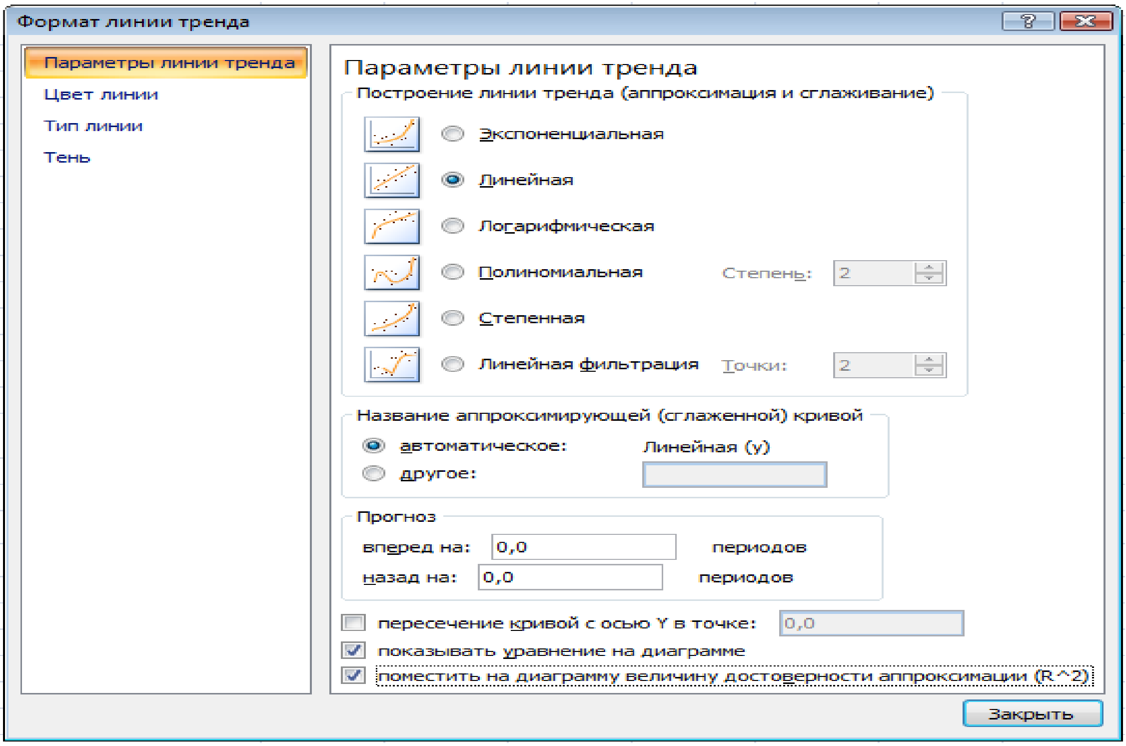

Рисунок 3.5 – Диалоговое окно выбора линии тренда

С помощью правой клавиши откроем контекстное меню (рисунок 3.5) и выберем одну из перечисленных линий трендов:

– Линейная;
– Логарифмическая;
– Полиномиальная;
– Степенная;
– Экспоненциальная;
– Линейная фильтрация (скользящая средняя).

После выбора одного из трендов, например, линейного – выберем и заполним вкладку Параметры диалогового окна (рисунок 3.5).

Можно выбрать название (назвать тренд самостоятельно) или оставить автоматически предлагаемое Excel; для прогноза согласно выбранной линии тренда на 5 лет вперед выберем соответствующее значение в диалоговом окне; для отображения на диаграмме уравнения тренда и коэффициента детерминации отметим соответствующие элементы вкладки Параметры (рисунок 3.6) Далее выберем OK.

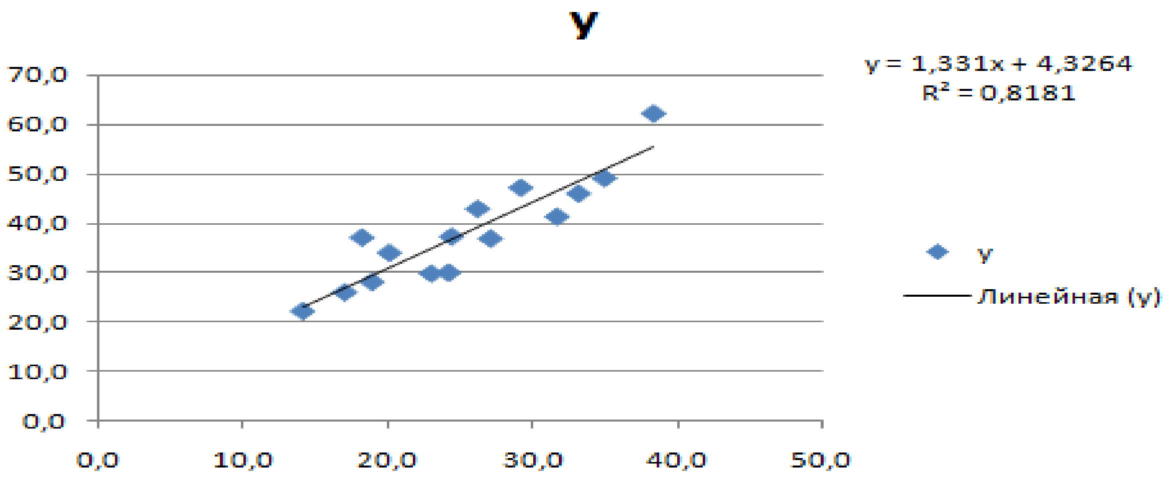

Рисунок 3.6 – График линейного уравнения

MSExcel позволяет проиллюстрировать основное свойство коэффициента корреляции – характеристики «степени линейности» наблюдаемой совокупности пар данных (рисунок 3.6).

Пример 3.2. Исследовать влияние продуктивности коров и производительности труда работников животноводства на себестоимость производства молока в сельскохозяйственных организациях.

Результативным признаком (Y) является себестоимость производства 1 ц молока.

Факторные признаки:

x_1 – среднегодовой надой молока на корову, характеризующий уровень продуктивности коров;

x_2 – прямые затраты труда на 1 ц молока, выражающий уровень трудоемкости производства и являющимся обратным показателем производительности труда.

Исходные данные представлены в таблице 3.3. Известны так же парные коэффициенты корреляции:

$$r_{yx_1} = -0{,}7905;\ r_{yx_2} = 0{,}6270;\ r_{x_1x_2} = -0{,}5812.$$

Требуется определить:

1) параметры множественного уравнения регрессии в натуральной и стандартизованной форме;

2) средние коэффициенты эластичности для каждого фактора;

3) коэффициенты частной и множественной корреляции;

4) общий и частные F-критерии Фишера.

Таблица 3.3 – Средние значения и колеблемость изучаемых признаков

Признак	Обозначение переменной	Среднее значение	Среднее квадратическое отклонение
Себестоимость 1 ц молока, руб.	Y	1823,36	369,54
Среднегодовой надой молока на корову, ц	X_1	59,85	14,75
Прямые затраты труда на 1 ц молока, чел.-ч	X_2	2,14	1,06

Решение

1. Линейное уравнение множественной регрессии в натуральной форме имеет вид:

$$y = b_0 + b_1 x_1 + b_2 x_2,$$

а в стандартизованной форме:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2}.$$

Найдем β-коэффициенты, используя парные коэффициенты корреляции.

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{-0{,}7905 - 0{,}6270 \cdot (-0{,}5812)}{1 - 0{,}5812^2} = -0{,}6435;$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0{,}6270 - (-0{,}7905) \cdot (-0{,}5812)}{1 - 0{,}5812^2} = 0{,}2530.$$

Линейное уравнение множественной регрессии в стандартизованном масштабе имеет вид:

$$t_y = -0{,}6435 t_{x_1} + 0{,}2530 t_{x_2}.$$

По абсолютной величине β-коэффициентов можно сделать вывод об относительной силе влияния факторов на изменение результативного признака.

Следовательно, на себестоимость производства молока более сильное влияние оказывает надой молока на корову, а влияние трудоемкости производства молока оказывает меньшее влияние.

Для определения параметров множественного уравнения регрессии в натуральной форме воспользуемся формулами:

$$b_j = \beta_j \frac{\sigma_y}{\sigma_{x_j}}\ , b_0 = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2. \quad (3.10)$$

$$b_1 = \beta_1 \frac{\sigma_y}{\sigma_{x_1}} = -0{,}6435 \cdot \frac{369{,}54}{14{,}75} = -16{,}1220;$$

$$b_2 = \beta_2 \frac{\sigma_y}{\sigma_{x_1}} = 0{,}253 \cdot \frac{369{,}54}{1{,}06} = 88{,}2015;$$

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 1823{,}36 - (-16{,}122)\cdot 59{,}85 - 88{,}2015\cdot 2{,}14 = \\ = 2599{,}5105.$$

Получим линейное уравнение множественной регрессии:

$$\hat{y} = 2599{,}5105 - 16{,}122\, x_1 + 88{,}2015\, x_2.$$

Коэффициенты множественной регрессии показывают, что при увеличении среднегодового надоя молока на корову на 1ц себестоимость молока в среднем снижается на 16,12 руб., а при увеличении прямых затрат труда на 1 ц молока на 1 чел.-ч себестоимость производства молока в среднем увеличивается на 88,20 руб.

2. Средние коэффициенты эластичности находятся по формуле:

$$\overline{Э}_{yx_j} = b_1 \frac{\overline{X}_j}{\overline{Y}}; \quad (3.11)$$

$$\overline{Э}_{yx_1} = b_1 \frac{\overline{X}_1}{\overline{Y}} = -16{,}122 \cdot \frac{59{,}85}{1823{,}36} = -0{,}529;$$

$$\overline{Э}_{yx_2} = b_2 \frac{\overline{X}_2}{\overline{Y}} = 88{,}2015 \cdot \frac{2{,}14}{1823{,}36} = 0{,}104.$$

Значит, при увеличении среднегодового надоя молока на корову на 1 %, себестоимость производства молока снижается в среднем на 0,529 %, при исключении влияния трудоемкости. Если увеличить трудоемкость производства молока на 1 %, то себестоимость молока в среднем увеличивается на 0,104 %, при исключении влияния продуктивности коров.

3. Коэффициенты частной корреляции определяются через парные коэффициенты корреляции по формулам:

$$r_{yx_1.x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)} = \frac{-0{,}7905 - 0{,}6270 \cdot (-0{,}5812)}{\sqrt{(1 - 0{,}627^2)(1 - 0{,}5812^2)}} = \\ = -0{,}6722;$$

$$r_{yx_2.x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)} = \frac{0{,}627 - (-0{,}7905) \cdot (-0{,}5812)}{\sqrt{(1 - 0{,}7905^2)(1 - 0{,}5812^2)}} = 0{,}3363;$$

$$r_{x_1x_2.y} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)} = \frac{-0{,}5812 - (-0{,}7905) \cdot (0{,}627)}{\sqrt{(1 - 0{,}7905^2)(1 - 0{,}627^2)}} = \\ = -0{,}1804.$$

Коэффициенты частной корреляции характеризуют тесноту связи между двумя переменными, при исключении влияния третьей переменной. Значит, связь между себестоимостью производства молока и продуктивностью коров обратная и тесная, исключив влияние трудоемкости. Связь между трудоемкостью и себестоимостью молока довольно слабая при исключении влияния продуктивности. Связь между факторами X_1 и X_2 очень слабая.

Коэффициент множественной корреляции находится по формуле:

$$R_{yx_1x_2} = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2}} = \sqrt{(-0{,}6435) \cdot (-0{,}7905) + 0{,}253 \cdot 0{,}627} = \\ = \sqrt{0{,}5087 + 0{,}1586} = \sqrt{0{,}6673} = 0{,}8169.$$

Величина коэффициента множественной корреляции показывает, что связь между Y, X_1 и X_2 очень тесная. Множественное уравнение регрессии объясняет 66,7 % ($0{,}8169^2 \times 100\%$) вариации себестоимости производства молока, в том числе влиянием продуктивности коров объясняется 50,9 % вариации себестоимости, а трудоемкостью производства – 15,8 %.

4. Оценим значимость уравнения регрессии и коэффициента R^2 с помощью F-критерия Фишера. Наблюдаемое или фактическое значение критерия находится по формуле:

$$F_{\text{набл}} = \frac{R^2_{yx_1x_2}}{1-R^2_{yx_1x_2}} : \frac{p}{n-p-1}. \qquad (3.12)$$

где p – число факторов в линейном уравнении регрессии;

n – число единиц наблюдения.

$$F_{\text{набл}} = \frac{0,6673}{1-0,6673} : \frac{2}{48-2-1} = 45,13.$$

При уровне значимости α = 0,05 и числе степеней свободы $k_1 = p = 2$, $.k_2 = n - p - 1 = 48 - 2 - 1 = 45$ по таблице значений F-критерия Фишера критическое значения $F_{\text{кр}}$ = 3,23. Сравниваем $F_{\text{набл}}$ с $F_{\text{кр}}$.

Так как $F_{\text{набл}} > F_{\text{кр}}$, то нулевую гипотезу о незначимости величины R^2 отклоним, т. е. уравнение множественной регрессии и R^2 статистически значимы.

В уравнении множественной регрессии не все факторы могут оказывать статистически существенное влияние на изменение результативного признака. Оценка значимости факторов в уравнении регрессии может быть дана с помощью частного F-критерия или t-критерия Стьюдента.

$$F_{\text{набл}\,x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1-R^2_{yx_1x_2}} \cdot \frac{n-p-1}{1} = \frac{0,6673 - 0,627^2}{1-0,6673} \cdot \frac{48-2-1}{1} = 37,08.$$

При $\alpha = 0,05, k_1 = 1, k_2 = 45, F_{\text{кр}} = 4,08$. Так как $F_{\text{набл}\,x_1} > F_{\text{кр}}$, то в уравнение регрессии целесообразно включение фактора X_1 после X_2. Фактор X_1 оказывает статистически значимое влияние на Y.

$$F_{\text{набл}\,x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1-R^2_{yx_1x_2}} \cdot \frac{n-k-1}{1} = \frac{0,6673 - 0,7905^2}{1-0,6673} \cdot \frac{48-2-1}{1} = 5,74.$$

$F_{кр} = 4{,}08$. Так как $F_{набл\,x_2} > F_{кр}$, то это свидетельствует о статистической значимости влияния фактора X_2 и целесообразности включения его в уравнение множественной регрессии. В данной задаче на уровень себестоимости производства молока статистически значимое влияние оказывает как продуктивность коров, так и трудоемкость производства продукции.

Пример 5.3. Исследовать влияние расхода кормов на 1 гол. и затрат труда на 1 ц на удой молока от одной коровы по сельскохозяйственным организациям региона.

Таблица 3.4 – Исходные данные для регрессионного анализа в MS Excel

№ п/п	Удой молока от одной коровы, ц/гол. (Y)	Расход кормов на 1 гол., ц корм. ед. (X_1)	Затраты труда на 1 ц, чел.-час (X_2)
1	37,8	44,9	3,02
2	60,7	69,3	6,87
3	36,8	22,7	2,55
4	41,4	21,6	4,82
5	40,0	26,9	4,55
6	42,3	29,5	1,92
7	42,6	61,2	3,29
8	45,2	59,9	6,81
9	49,6	60,1	6,77
10	57,2	71,7	5,84
11	41,7	48,5	3,16
12	47,5	50,3	5,93
13	30,4	21,4	1,54
14	38,0	40,1	2,83
15	54,5	55,0	5,97
16	44,4	59,2	4,99
17	38,2	40,1	1,58
18	38,5	45,3	5,02
19	35,9	29,1	4,04
20	35,7	21,3	2,93

Результативным признаком (Y) является удой молока от одной коровы, ц/гол.

Факторные признаки:

X_1 – расход кормов на 1 гол., ц корм. ед.;

X_2 – затраты труда на 1 ц, чел.-час.

Определить:

1) параметры множественного уравнения регрессии в натуральной и стандартизованной форме;

2) средние коэффициенты эластичности для каждого фактора;

3) коэффициенты частной и множественной корреляции;

4) общий и частные F-критерии Фишера.

Решение

Связь между результативным признаком Y и факторами X_1 и X_2 выражается множественным линейным уравнением регрессии, которое имеет вид: $\hat{y} = b_0 + b_1x_1 + b_2x_2$.

Рассмотрим применение пакета анализа данных в *Excel MS Office* для решения задачи. Исходные данные введем на листе *MS Excel* в виде, представленном таблице 3.4.

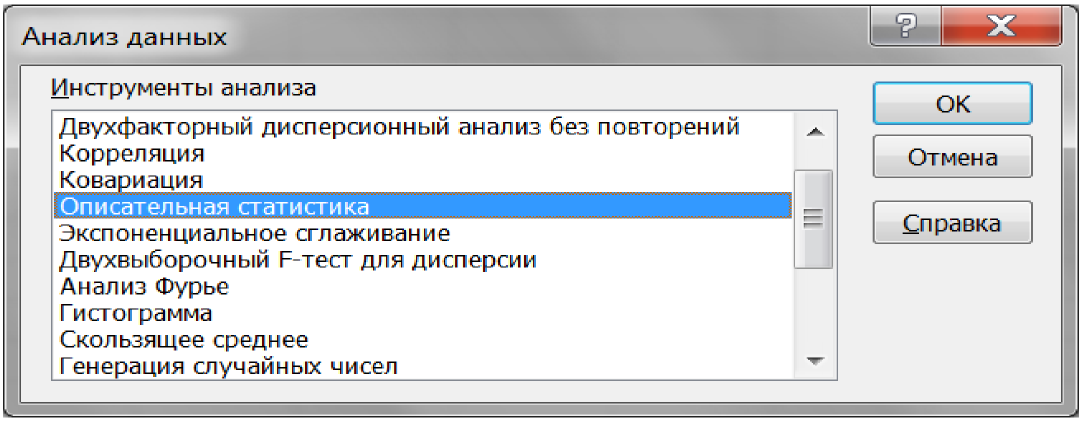

Рисунок 3.7 – Диалоговое окно пакета анализа данных

Для проведения анализа предварительно установим пакет анализа, выполнив последовательно действия: кнопка *Office – Параметры Excel – Надстройки – Пакет анализа – Перейти* (выделим в окне доступных надстроек *Пакет анализа*). После этого на вкладке *Данные* ленты появится инструмент *Пакет анализа*.

Выберем в *Пакете анализа* инструмент *Описательная статистика* (рисунок 3.7) и заполним параметры диалогового окна (рисунок 3.8).

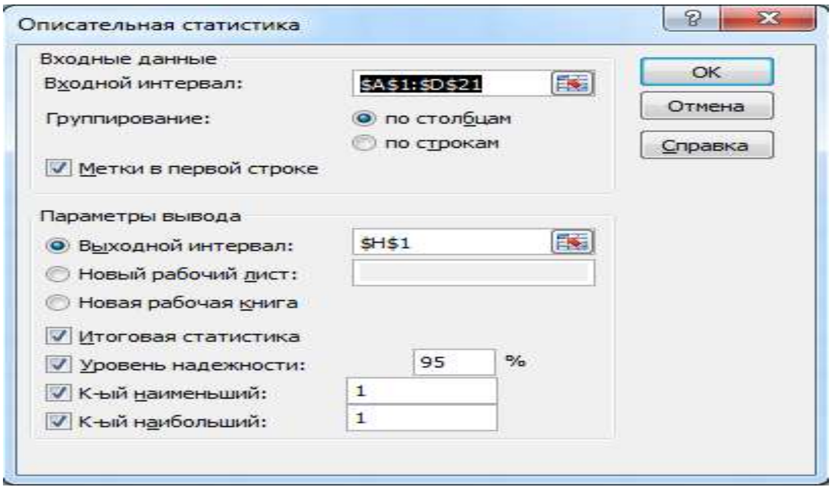

Рисунок 3.8 – Диалоговое окно описательной статистики

В результате будут рассчитаны обобщающие характеристики по каждому признаку (таблица 3.5).

Таблица 3.5 – Обобщающие характеристики исследуемых признаков

Показатель	*У*	*Х*₁	*Х*₂	Принятые обозначения
Среднее значение	42,921	43,908	4,222	$\overline{X} = \sum x_i n_i / n$
Стандартная ошибка	1,722	3,760	0,397	$s_{\overline{X}} = s / \sqrt{n}$
Медиана	41,548	45,108	4,295	M_e
Мода	н/д	н/д	н/д	M_o
Стандартное отклонение	7,700	16,816	1,774	s
Дисперсия выборки	59,289	282,768	3,148	$s^2 = \sum (x_i - \overline{X})^2 n_i /(n-1)$
Эксцесс	0,471	–1,321	–1,289	$Ex = \sum ((x_i - \overline{X}) / S)^4 n_i / n - 3$
Асимметричность	0,909	0,022	0,053	$Sk = \sum ((x_i - \overline{X}) / S)^3 n_i / n$
Интервал	30,3	50,426	5,33	$W = x_{max} - x_{min}$
Минимум	30,4	21,3	1,54	x_{min}
Максимум	60,7	71,7	6,87	x_{max}
Сумма	858,42	878,2	84,43	Σx_i
Счет	20	20	20	$n = \Sigma n_i$
Наибольший (1)	60,7	71,7	6,87	–
Наименьший (1)	30,4	21,3	1,54	–
Уровень надежности (95,0 %)	3,61	7,87	0,83	$\Delta = t_{\alpha;n-1} s_{\overline{X}}$

Данные таблицы 3.5 показывают, что по совокупности предприятий средний удой от одной коровы составил 42,9 ц/гол. и в среднем между предприятиями продуктивность коров колеблется в границах 42,9 ± 7,7 ц/гол., т. е. от 35,2 до 50,6 ц/гол. Коэффициент вариации составил 17,9 %, что свидетельствует о больших различиях в удое молока от одной коровы между предприятиями. По значению медианы видно, что половина предприятий имеет размер удоя до 41,5 ц/гол., а половина более. Распределение предприятий по данному признаку является несимметричными (*Ка* = 0,909) и островершинным (Э = – 0,471). Наименьшее значение удоя молока от одной коровы составило 30,4 ц/гол., а наибольшее – 60,7 ц/гол.

Аналогичные выводы можно сделать и по факторным признакам X_1 *и* X_2.

Для нахождения парных коэффициентов корреляции применим инструмент пакета анализа *Корреляция*, для этого заполним параметры диалогового окна как на рисунке 3.9.

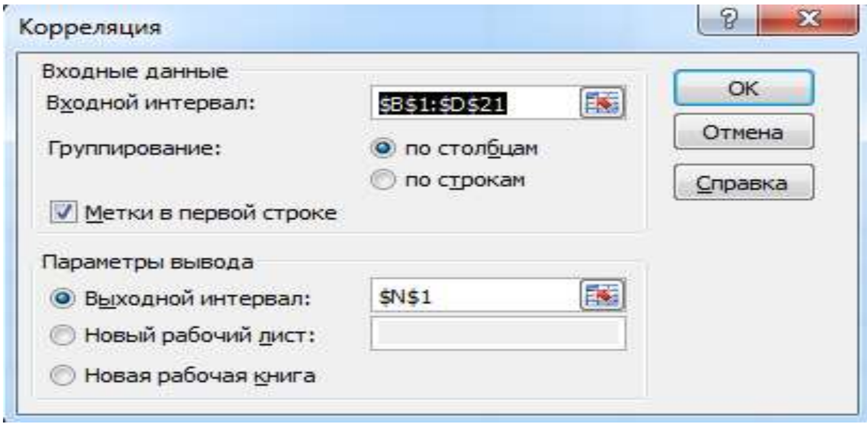

Рисунок 3.9 – Диалоговое окно «Корреляция»

В результате будет получена матрица парных коэффициентов корреляции между всеми изучаемыми переменными (таблица 3.6).

Таблица 3.6 – Парные коэффициенты корреляции между признаками

	У	X_1	X_2
У	1	0,7963	0,7692
X_1	0,7963	1	0,6586
X_2	0,7692	0,6586	1

Значит: r_{yx_1} = 0,7963; r_{yx_2}= 0,7692; $r_{x_1x_2}$= 0,6586.

Парные коэффициенты корреляции показывают, что связь между удоем молока от одной коровы, расходом концентрированных кормов на 1 гол. и затратами труда на 1 ц довольно тесная, а между факторными признаками X_1 и X_2 – средняя.

Линейное уравнение множественной регрессии в натуральной форме имеет вид:

$$y = b_o + b_1 x_1 + b_2 x_2.$$

Найдем параметры этого уравнения, используя инструмент *Пакета анализа – Регрессия.* Заполним параметры диалогового окна (рисунок 3.10).

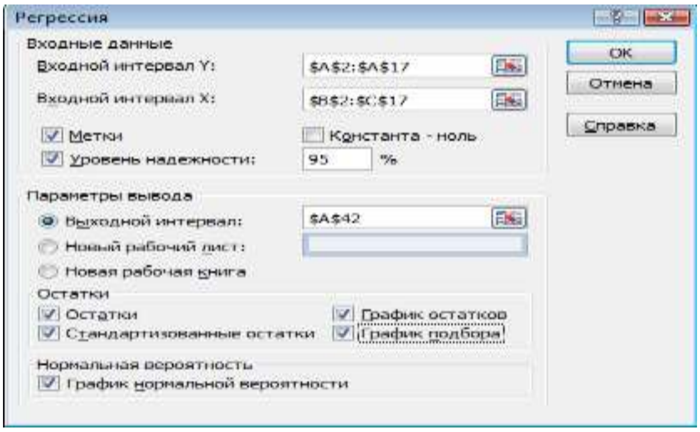

Рисунок 3.10 – Диалоговое окно «Регрессия»

Согласно рисунку 3.11 линейное уравнение множественной регрессии имеет вид:

$$y = 24{,}699 + 0{,}234x_1 + 1{,}877x_2.$$

Коэффициенты множественной регрессии показывают, что при увеличении расхода концентрированных кормов на 1 гол. на 1 ц корм ед. удой молока от одной коровы в среднем увеличивается на 0,234 ц/гол. (при исключении влияния второго фактора X_2), а при росте затрат труда на 1 ц на 1 чел.-ч он в среднем возрастает на 1,877 ц/гол.

Найдем β-коэффициенты, используя их связь с коэффициентами b_j уравнения регрессии в нормальной форме:

$$\beta_j = b_j \frac{\sigma_{x_j}}{\sigma_y}, \tag{3.13}$$

$$\beta_1 = 0{,}234 \cdot \frac{16{,}816}{7{,}7} = 0{,}512,$$
$$\beta_2 = 1{,}877 \cdot \frac{1{,}774}{7{,}7} = 0{,}432.$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,8602
R-квадрат	0,7399
Нормированный R-квадрат	0,7093
Стандартная ошибка	4,1536
Наблюдения	20

Дисперсионный анализ						
	df	*SS*	*MS*	*F*	*Значимость F*	
Регрессия	2	834,4994099	417,249705	24,18487	1,0669E-05	
Остаток	17	293,2925901	17,2525053			
Итого	19	1127,792				
Коэффициенты		*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	24,699	2,78480	8,8693	8,718E-08	18,8239	30,5747
X1	0,2345	0,0753	3,1125	0,0063	0,0756	0,3934
X2	1,8770	0,71356	2,6304	0,0175	0,3715	3,3825

Рисунок 3.11 – Вывод итогов регрессионного анализа

β-коэффициенты, можно также найти с помощью парных коэффициентов корреляции по формулам:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} = \frac{0{,}7963 - 0{,}7692 \cdot 0{,}6568}{1 - 0{,}6568^2} = 0{,}512,$$
$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} = \frac{0{,}7692 - 0{,}7963 \cdot 0{,}6568}{1 - 0{,}6568^2} = 0{,}432.$$

Линейное уравнение множественной регрессии в стандартизованном масштабе имеет вид:

$$t_y = 0{,}512 t_{x_1} + 0{,}432\, t_{x_2}.$$

По абсолютной величине β-коэффициентов можно сделать вывод об относительной силе влияния факторов на изменение результативного признака. Видно, что на продуктивность коров большее влияние оказывает расход концентрированных кормов и меньшее – затраты труда.

2. Средние коэффициенты эластичности находятся по формуле:

$$Эyx_1 = b_j \cdot \frac{\bar{x}_j}{\bar{y}}, \qquad (3.14)$$

$$Эyx_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 0{,}234 \cdot \frac{43{,}908}{42{,}921} = 0{,}239,$$

$$Эyx_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 1{,}877 \cdot \frac{4{,}222}{42{,}921} = 0{,}185.$$

Значит, при увеличении расхода кормов на 1 % удой молока увеличивается в среднем на 0,239 %, исключив влияние второго фактора. Если увеличить затраты труда на 1 ц на 1 %, то удой молока от одной коровы в среднем возрастет на 0,185 %, исключив влияние первого фактора.

2. Коэффициенты частной корреляции определяются через парные коэффициенты корреляции по формулам:

$$r_{yx_1\text{-}x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}} = \frac{0{,}7963 - 0{,}7692 \cdot 0{,}6568}{\sqrt{(1 - 0{,}7692^2)(1 - 0{,}6568^2)}} = 0{,}604;$$

$$r_{yx_2\text{-}x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}} = \frac{0{,}7692 - 0{,}7963 \cdot 0{,}6568}{\sqrt{(1 - 0{,}7963^2)(1 - 0{,}6568^2)}} = 0{,}540;$$

$$r_{x_1x_2\text{-}y} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{yx_2})}} = \frac{0{,}6568 - 0{,}7963 \cdot 0{,}7692}{\sqrt{(1 - 0{,}7963^2)(1 - 0{,}7692^2)}} = 0{,}115.$$

Коэффициенты частной корреляции характеризуют тесноту связи между двумя переменными, исключив влияние третьей переменной. Значит, связь между расходами концентрированных кормов и удоем молока довольно тесная и прямая, между затратами труда на 1 ц и продуктивностью коров прямая и средняя.

А связь между факторами x_1 и x_2 слабая и также прямая.

Коэффициент множественной корреляции находится по формуле:

$$R_{yx_1x_2} = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2}} = \sqrt{0{,}512 \cdot 0{,}7963 + 0{,}432 \cdot 0{,}7692} =$$

$$\sqrt{0{,}408 + 0{,}332} = \sqrt{0{,}74} = 0{,}86.$$

Величина коэффициента множественной корреляции показывает, что связь между продуктивностью коров и обоими факторами очень тесная, причем 73,9 % вариации продуктивности объясняется влиянием обоих факторов, из которой на долю первого приходится 36,5 % вариации, а второго – 29,2 %.

3. Оценим значимость уравнения регрессии и множественного коэффициента детерминации R^2 с помощью F-критерия Фишера.

Фактически рассматривается нулевая гипотеза $H_0: R^2 = 0, (b_1 = b_2 = 0)$ и альтернативная гипотеза $H_1: R^2 \neq 0, (b_1 \neq 0, b_2 \neq 0)$.

Наблюдаемое значение критерия находится по формуле:

$$F_н = \frac{R^2_{yx_1x_2}}{1 - R^2_{yx_1x_2}} : \frac{m}{n - m - 1}, \qquad (3.15)$$

где m – число факторов в линейном уравнении регрессии;

n – число единиц наблюдения.

$$F_н = \frac{0{,}737}{1 - 0{,}737} : \frac{2}{20 - 2 - 1} = 24{,}19.$$

При уровне значимости α = 0,05 и числе степеней свободы $k_1 = m = 2$, $k_2 = n - m - 1 = 20 - 2 - 1 = 17$, по таблице значений F-критерия Фишера критическое значения составляет 3,59, т. е. $F_{кр} = 3{,}59$.

Сравниваем $F_н$ с $F_{кр}$. Так как $F_н > F_{кр}$, то нулевую гипотезу о незначимости величины R^2 отклоняем, т. е. уравнение множественной регрессии и R^2 статистически значимы.

В уравнении множественной регрессии не все факторы могут оказывать статистически существенное влияние на изменение результативного признака. Оценка значимости факторов в уравнении регрессии может быть дана с помощью частного F-критерия Фишера или t-критерия Стьюдента.

$$F_{нx_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1} = \frac{0{,}74 - 0{,}5917}{1 - 0{,}74} \cdot \frac{20 - 2 - 1}{1} = 9{,}70.$$

При α = 0,05, k_1 = 1, k_2 = 17, $F_{кр}$ = 4,45. Так как $F_{нx_1} > F_{кр}$, то в уравнение регрессии целесообразно включение фактора X_1 после X_2. Фактор X_1 оказывает статистически значимое влияние на $У$.

$$F_{нx_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1} = \frac{0,737-0,6341}{1-0,74} \cdot \frac{20-2-1}{1} = 6,92.$$

В этом случае также наблюдаемое значение критерия Фишера больше критического, это свидетельствует о статистической значимости влияния фактора X_2 и целесообразности включения его в уравнение множественной регрессии. В данной задаче на выручку от реализации продукции статистически значимое влияние оказывают оба фактора. Небольшие расхождения в результатах расчетов в компьютерном и ручном вариантах расчетов обусловлено округлением расчетных значений.

Для определения тесноты связи между двумя признаками, измеренными в порядковых шкалах, применяются менее точные, но более простые по расчету непараметрические показатели, в частности **коэффициенты корреляции рангов** (или ранговые коэффициенты корреляции) Спирмена (ρ) и Кендалла (τ). Оба показателя основаны на корреляции не самих значений изучаемых признаков, а их рангов.

Ранг – это порядковый номер, присваиваемый каждому индивидуальному значению x и y (отдельно) в ранжированном ряду.

Пример 3.4. С помощью коэффициентов корреляции рангов Спирмена и Кендалла, измерить тесноту связи между уровнем жирности молока и содержанием белка в молоке дочерей двух проверяемых быков – производителей (таблицы 3.7 и 3.8).

Решение

Расчет показателей для дочерей быка Буцефала:

1) Коэффициента корреляции рангов Спирмена:

$$\rho = 1 - \frac{6\sum_{i=1}^{10} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 84}{10(100-1)} = +0,491.$$

Таблица 3.7 – Ранговое распределение коров – дочерей двух быков

Номер коровы	Жир, (x)		Белок, (y)		Разность рангов (x-y)	Квадрат разности рангов $(x\text{-}y)^2$	Число рангов R_{yi}	
	%	Ранг, (x)	%	Ранг, (y)			большего порядка	меньшего порядка
	Дочери быка БУЦЕФАЛА (1)							
1	3,6	1	3,1	2,5	–1,5	2,25	7	1
2	3,5	2	3,1	6,5	–4,5	20,25	3	3
3	3,4	3,5	3,1	1	+2,5	6,25	7	0
4	3,4	3,5	3,1	6,5	–3,0	9,00	3	3
5	3,3	5,5	3,0	8,5	–3,0	9,00	1	3
6	3,3	5,5	3,2	4,5	+1,0	1,00	2	2
7	3,2	7	3,3	2,5	+4,5	20,25	3	0
8	3,1	8,5	3,4	4,5	+4,0	16,00	2	0
9	3,1	8,5	3,0	8,5	0	0,00	1	0
10	3,1	10	2,9	10	0	0,00	–	–
СУММА	33,0	–	31,2	–	–	84,00	P = 29	Q = 12

Таблица 3.8 – Ранговое распределение коров – дочерей двух быков

Номер коровы	Жир, (x)		Белок, (y)		Разность рангов (x-y)	Квадрат разности рангов $(x\text{-}y)^2$	Число рангов R_{yi}	
	%	Ранг, (x)	%	Ранг, (y)			большего порядка	меньшего порядка
	Дочери быка ЦЕЗАРЯ (2)							
1	3,9	1	3,6	3	–2	4	7	2
2	3,8	2,5	3,7	1,5	+1	1	7	0
3	3,8	2,5	3,7	1,5	+1	1	7	0
4	3,7	4,0	3,5	4,5	–0,5	0,25	5	0
5	3,6	5,5	3,5	4,5	+1	1	5	0
6	3,6	5,5	3,4	6,5	–1	1	3	0
7	3,4	7,5	3,4	6,5	+1	1	3	0
8	3,4	7,5	3,3	8	–0,5	0,25	2	0
9	3,3	9	3,2	9	0	0	1	0
10	3,2	10	3,1	10	0	0	–	–
СУММА	35,7	–	34,4	–	–	9,5	P = 40	Q = 2

2) Коэффициент корреляции рангов Кендалла

$$\tau = \frac{2S}{n(n-1)} = \frac{2 \cdot (29 - 12)}{10 \cdot (10 - 1)} = +0{,}378.$$

3) Рассчитаем средние значения

$$\overline{x_x} = \frac{\sum x_1}{n_1} = \frac{33{,}0}{10} = 3{,}30\ \%,$$
$$\overline{y_y} = \frac{\sum y_1}{n_1} = \frac{31{,}2}{10} = 3{,}12\ \%.$$

4) Проверим, существенна ли положительная средняя корреляционная связь между уровнем жирности молока и содержанием белка у дочерей Буцефала, для этого (при $n \leq 30$) используем t-статистику Стьюдента с $k = n - 2$ степенями свободы:

$$t_{\text{н}} = |r| \cdot \sqrt{\frac{n-2}{1-r_s^2}}.$$

Нулевая гипотеза – коэффициент корреляции не является статистически значимым ($\text{H}_0 : r_s = 0$). Альтернативная гипотеза – существует положительная корреляционная зависимость ($\text{H}_1 : r_s > 0$). При уровне значимости α = 0,05 для односторонней (правосторонней) критической области (приложение Г):

$$t_{\text{кр}} = t(\alpha = 0{,}05; k = 8) = 1{,}86,$$
$$t_{\text{н}} = |0{,}491| \cdot \sqrt{\frac{10-2}{1-0{,}491^2}} = 1{,}59.$$

Так как $t_{\text{набл}} < t_{\text{кр}}$, то связь между уровнем жирности молока и содержанием белка у дочерей Буцефала не является статистически значимой при 5 % уровне значимости.

Расчет показателей для дочерей быка Цезаря:

1) Коэффициента корреляции рангов Спирмена

$$\rho = 1 - \frac{6\sum_{i=1}^{10} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 9{,}5}{10(100-1)} = +0{,}942.$$

2) Коэффициент корреляции рангов Кендалла

$$\tau = \frac{2S}{n(n-1)} = \frac{2 \cdot (40-2)}{10 \cdot (10-1)} = +0{,}844.$$

3) Рассчитаем средние значения

$$\overline{x_x} = \frac{\sum x_1}{n_1} = \frac{34{,}8}{10} = 3{,}48\ \%,$$
$$\overline{y_y} = \frac{\sum y_1}{n_1} = \frac{33{,}4}{10} = 3{,}34\ \%.$$

4) Проверим, существенна ли положительная высокая корреляционная связь между уровнем жирности молока и содержанием белка у дочерей Цезаря, для этого (при $n \leq 30$) используем t-статистику Стьюдента с $k = n - 2$ степенями свободы:

$$t_{\text{н}} = |r| \cdot \sqrt{\frac{n-2}{1-r_s^2}}.$$

Нулевая гипотеза – коэффициент корреляции не является статистически значимым ($\text{H}_0 : r_s = 0$). Альтернативная гипотеза – существует положительная корреляционная зависимость ($\text{H}_1 : r_s > 0$). При уровне значимости α = 0,05 для односторонней (правосторонней) критической области (приложение Г):

$$t_{\text{кр}} = t(\alpha = 0{,}05; k = 8) = 1{,}86,$$
$$t_{\text{н}} = |0{,}942| \cdot \sqrt{\frac{10-2}{1-0{,}942^2}} = 7{,}93.$$

Так как $t_{\text{набл}} > t_{\text{кр}}$, то связь между уровнем жирности молока и содержанием белка у дочерей Цезаря является статистически значимой при 5% уровне значимости.

Вывод: связь жирности молока и содержания белка в нем у дочерей двух быков различная. Она оказалась очень высокой у дочерей быка Цезаря ($r_s = +0{,}942$) и умеренной у потомства Буцефала ($r_s = +0{,}491$). Значит селекционный эффект на одновременное повышение белка и жира в молоке дочерей наиболее высок у потомков быка Цезаря. Высокий коэффициент корреляции свидетельствует, что в потомстве быка Цезаря косвенная селекция на белок в молоке по показателям жирности молока проявляется сильнее. Дочери быка Буцефала по содержанию жира и содержанию белка в молоке, а также по степени связи между этими признаками представляют меньшую селекционную ценность.

При исследовании тесноты связи между **качественными признаками** строят **таблицы сопряженности**.

Если каждый из двух качественных признаков принимает только альтернативные значения, то таблица сопряженности имеет вид:

Таблица 3.9 – Таблица сопряженности

Признак А	Признак В	
	да	нет
Да	a	b
Нет	c	d

Каждая из клеток данной таблицы соответствует известной альтернативе того или другого признака. Для оценки тесноты связи рассчитывают **коэффициенты ассоциации (*К*$_a$) и контингенции (*К*$_к$)**:

$$K_a = \frac{ad - bc}{ad + bc}, \qquad (3.16),$$

$$K_к = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \qquad (3.17)$$

Данные коэффициенты изменяются от –1 до +1 , причем коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается значимой и подтвержденной, если $|K_a| \geq 0{,}5$, а $|K_к| \geq 0{,}3$.

Для определения тесноты связи применяют **коэффициенты взаимной сопряженности Пирсона и Чупрова**.

Коэффициент Пирсона исчисляется по формуле:

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}, \qquad (3.18)$$

где φ^2 – показатель взаимной сопряженности (φ – фи).

$$\varphi^2 = \sum_{i=1}^{s} \sum_{j=1}^{p} \frac{f_{ij}^2}{n_i m_j} - 1, \qquad (3.19)$$

где s – число категорий (значений) признака А;

p – число категорий признака В.

Коэффициент Чупрова исчисляется по формуле:

$$K_Ч = \sqrt{\frac{\varphi^2}{\sqrt{(s-1)(p-1)}}}\ . \qquad (3.20)$$

Чем ближе C и $Kч$ к 1, тем связь теснее. Коэффициент взаимной сопряженности Чупрова более точен, поскольку учитывает число значений по каждому признаку.

Пример 3.6. По данным таблицы 3.10 с помощью коэффициентов взаимной сопряженности Пирсона и Чупрова оценить наличие связи между окрасочными морфами сизого голубя и средой обитания. В процессе статистического наблюдения голубей подсчитывали на разных площадках приблизительно одинаковой протяженности. По окрасу оперения были выделены четыре градации. Две из них (сизая и черночеканная) – типичны для данного вида, тогда как наличие непигментированных и красных перьев (пегая и красная окраска соответственно) для диких представителей данного вида в целом несвойственно. В ходе последующего анализа пегие и красные окрасочные морфы в связи с их редкостью объединены в одну категорию «нетипичных фенов».

Таблица 3.10 – Соотношение окрасочных морф сизого голубя

Пробная площадка	Окрасочные морфы, гол.			Всего
	сизые	чёрночеканные	нетипичные	
1	43	66	44	153
2	47	75	25	147
3	52	19	3	74
Итого:	142	160	72	374

Решение

1. Рассчитаем показатель взаимной сопряженности φ^2:

$$\xi_1 = \left(\frac{43^2}{142} + \frac{66^2}{160} + \frac{44^2}{72}\right) : 153 = 0{,}437; \xi_2 = \left(\frac{47^2}{142} + \frac{75^2}{160} + \frac{25^2}{72}\right) : 147 = 0{,}406;$$

$$\xi_3 = \left(\frac{52^2}{142} + \frac{19^2}{160} + \frac{3^2}{72}\right) : 74 = 0{,}285$$

Тогда $\varphi^2 = \xi_1 + \xi_2 + \xi_3 - 1 = 0{,}437 + 0{,}406 + 0{,}285 - 1 = 0{,}129$

2. Определим коэффициент Пирсона:

$$C = \sqrt{\frac{0{,}129}{1 + 0{,}129}} = 0{,}338.$$

3. Вычислим коэффициент Чупрова:

$$К_ч = \sqrt{\frac{0{,}129}{\sqrt{(3-1)(3-1)}}} = 0{,}254.$$

Вывод: следовательно, между окраской морф сизого голубя и средой обитания существует средняя, приближенная к слабой связь.

Задача 3.1. По 20 сельскохозяйственным организациям имеются данные по расходу кормов на корову и надою молока от коровы.

№ п/п	Надой молока на корову, ц	Расход кормов на корову, ц корм. ед.	№ п/п	Надой молока на корову, ц	Расход кормов на корову, ц корм. ед.
1	50,1	60,9	11	75,9	80,0
2	83,8	76,9	12	71,0	76,3
3	63,7	67,0	13	62,2	73,2
4	72,1	62,4	14	69,1	65,2
5	46,5	41,5	15	63,5	51,7
6	54,6	64,8	16	47,9	65,1
7	40,7	45,8	17	68,0	61,4
8	71,7	55,6	18	51,6	60,6
9	58,6	56,2	19	39,9	35,8
10	46,2	43,5	20	55,7	58,2

На основании имеющихся данных определить параметры линейного уравнения регрессии между уровнем кормления и продуктивностью коров, рассчитать коэффициенты корреляции и детерминации. Оценить существенность величины коэффициентов корреляции и регрессии при уровне значимости 0,05.

Задача 3.2. Получены следующие данные о продолжительности беременности у кроликов породы шиншилла при различных размерах помета (число крольчат в помете x и длительность беременности y):

Х 2; 8; 3; 5; 7; 8; 4; 8; 3; 4; 4; 8; 8; 5; 7; 6; 6; 5; 6; 6; 5; 7; 8; 10; 6; 7; 6; 7; 6; 5; 10; 7; 8; 8; 6; 5; 6; 5; 6; 8; 6; 5 9; 7; 6; 5; 9; 5; 3; 4; 7; 8; 9; 5; 6; 2; 2.

У 32; 30; 31; 31; 31; 32; 31; 31; 32; 33; 32; 31; 31; 31; 31; 30; 31; 32; 32; 32; 32; 31; 32; 31; 31; 30; 31; 31; 32; 31; 30; 32; 32; 31; 31; 31; 32; 30; 31; 31; 32; 32; 31; 30; 32; 31; 31; 31; 32; 32; 31; 31; 31; 31; 31; 32; 33.

Есть ли корреляция между длительностью плодоношения и размерами помета?

Задача 3.3. По данным приложения А по 20 организациям требуется:

1. Построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии.
2. Рассчитать параметры уравнения регрессии методом наименьших квадратов.
3. Оценить качество уравнения с помощью средней ошибки аппроксимации.
4. Найти коэффициент эластичности.
5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.
6. Оценить значимость коэффициентов корреляции и регрессии по t-критерию Стьюдента при уровне значимости $\alpha = 0{,}05$.
7. Охарактеризовать статистическую надежность результатов регрессионного анализа с использованием критерия F – Фишера при уровне значимости $\alpha = 0{,}05$.
8. Определить прогнозное значение результативного признака, если возможное значение факторного признака составит 1.2 от его среднего уровня по совокупности.

Задача 3.4. Изучить зависимость и выявить влияние факторов на изменение результативных признаков по следующим парам признаков:

1) Производственная себестоимость производства молока (руб.) и годовой надой молока на среднегодовую корову, кг;

2) Производственная себестоимость производства молока (руб.) и прямые затраты труда на 1 ц, чел.-ч;

3) Производственная себестоимость производства молока (руб.) и затраты по оплате труда на 1 ц, руб.;

4) Производственная себестоимость производства молока (руб.) и затраты на корма на 1 ц молока, руб.

5) Коммерческая себестоимость производства молока (руб.) и годовой надой молока на среднегодовую корову, кг;

6) Коммерческая себестоимость производства молока (руб.) и прямые затраты труда на 1 ц, чел.-ч;

7) Коммерческая себестоимость производства молока (руб.) и затраты по оплате труда на 1 ц, руб.;

8) Коммерческая себестоимость производства молока (руб.) и затраты на корма на 1 ц молока, руб.

Задача 3.5. По данным, представленным в приложении А, изучить влияние совокупности факторов на производственную и коммерческую себестоимость производства молока (y_1 и y_2) в сельскохозяйственных предприятиях северной зоны Краснодарского края за 2011 г.

1. С помощью инструмента анализа данных *Описательная статистика* рассчитать обобщающие характеристики вариационных рядов, написав выводы по каждой переменной.

2. С помощью инструмента анализа данных *Корреляция* построить матрицу парных коэффициентов корреляции. Оценить наличие или отсутствие мультиколлинеарности между факторами.

3. С помощью инструмента анализа данных *Регрессия* рассчитать параметры множественного уравнения регрессии с включением всех факторов.

4. Рассчитать частные коэффициенты корреляции, эластичности и стандартизованные коэффициенты регрессии.

5. Оценить значимость множественного уравнения регрессии и множественных коэффициентов регрессии с помощью критериев Фишера и Стьюдента.

6. Определить среднюю ошибку аппроксимации.

7. Провести последовательный отсев статистически не значимых факторов и получить модель себестоимости производства молока со всеми статистически значимо влияющими факторами. Рассчитать частные коэффициенты корреляции, эластичности и стандартизованные коэффициенты регрессии. Оценить значимость множественного уравнения регрессии и множественных коэффициентов регрессии с помощью критериев Фишера и Стьюдента. Определить среднюю ошибку аппроксимации.

8. Написать выводы по результатам множественного регрессионного анализа себестоимости производства молока.

Задача 3.6. По данным 41 сельскохозяйственного предприятия северной зоны Краснодарского края изучается зависимость продуктивности коров от влияющих на нее факторов.

Таблица 3.11 – Статистические характеристики выборочной совокупности сельскохозяйственных предприятий

Показатель	y	x_1	x_2	x_3	x_4	x_5	x_6
Среднее значение	5383,5	70,04	31,74	130,58	961,34	135,34	70,68
Стандартная ошибка	228,24	2,83	1,51	6,33	100,29	9,23	2,61
Медиана	5178,91	68,85	31,61	127,69	736,00	124,38	74,18
Среднее квадратическое отклонение	1461,47	18,15	9,66	40,54	642,16	59,09	16,73
Дисперсия выборки	2135883	329,27	93,29	1643,71	412372,98	3491,44	279,95
Эксцесс	0,480	3,700	–0,146	–0,432	1,753	1,067	0,160
Асимметричность	0,739	1,134	0,316	0,452	1,486	0,973	–0,867
Интервал	6936,9	98,9	43,7	154,5	2763,0	267,7	66,7
Минимум	2718,30	39,27	14,07	60	229	39,217	28,446
Максимум	9655,23	138,15	57,74	214,5	2992	306,921	95,167
Сумма	220722	2872	1302	5354	39415	5549	2898

Зависимая переменная (y) – годовой надой молока на среднегодовую корову, кг.

Объясняющие переменные: x_1 – производственные затраты на среднегодовую корову, тыс. руб.; x_2 – затраты на корма на среднегодовую корову, тыс. руб.; x_3– прямые затраты труда на среднегодовую корову, чел.-ч; x_4– среднегодовое поголовье коров на предприятие, гол.; x_5 – затраты по оплате труда на 1 чел.-ч, руб.; x_6 – доля молока в выручке от реализации продукции животноводства, %.

В таблицах 3.11 и 3.12 приведены статистические характеристики по изучаемой совокупности предприятий и парные коэффициенты корреляции между переменными.

Таблица 3.12 – Парные коэффициенты корреляции

	y	x_1	x_2	x_3	x_4	x_5	x_6
y	1						
x_1	0,8334	1					
x_2	0,6471	0,8165	1				
x_3	–0,1378	–0,1519	–0,0285	1			
x_4	0,2858	0,2166	0,1392	0,0212	1		
x_5	0,3106	0,3061	0,0705	–0,4933	0,0920	1	
x_6	0,2418	0,2456	0,3239	–0,1536	–0,3065	–0,0968	1

Варианты заданий к задаче 3.4

Вариант	1	2	3	4	5	6
Переменные	$y; x_1;x_4$	$y; x_1; x_5$	$y; x_1 ; x_6$	$y; x_2; x_3$	$y; x_2; x_4$	$y; x_2; x_5$
Вариант	7	8	9	10	11	12
Переменные	$y; x_2 ; x_6$	$y; x_4 ; x_5$	$y; x_4 ;x_6$	$y; x_5 ; x_6$	$y; x_3 ; x_4$	$y; x_3 ; x_5$

1. По одному варианту составить матрицу парных коэффициентов корреляции между тремя переменными.
2. Определить параметры множественного уравнения регрессии в стандартизированной и естественной форме.
3. Рассчитать частные коэффициенты эластичности.
4. Рассчитать частные и множественный коэффициенты корреляции и детерминации.
5. Оценить значимость множественного уравнения регрессии с помощью F-критерия Фишера, для чего составить таблицу дисперсионного анализа.
6. С помощью частных F-критериев Фишера оценить целесообразность включения фактора x_1 после x_2 и фактора x_2 после x_1.
7. Оценить значимость множественных коэффициентов регрессии с помощью t-критерия Стьюдента.
8. Написать выводы по результатам расчетов.

Задача 3.7. По данным, приведенным в приложении А, изучить влияние факторов x_2, x_3, x_4, x_5, x_6 на годовой удой на среднегодовую корову (y).

1. Определить параметры множественного уравнения регрессии с включением в линейную модель всех факторов.
2. Оценить значимость множественных коэффициентов регрессии с помощью t-критерия Стьюдента.
3. Провести последовательный отсев статистически не значимых факторов.
4. По уравнению со всеми значимо влияющими факторами рассчитать частные и множественный коэффициенты корреляции и детерминации.
5. Оценить значимость множественного уравнения регрессии с помощью F-критерия Фишера, для чего составить таблицу дисперсионного анализа.
6. Написать выводы по результатам расчетов.

Задача 3.8. Используя данные приложения А по одному варианту определить:

1. Параметры линейного множественного уравнения регрессии в натуральной и стандартизованной форме;
2. Средние коэффициенты эластичности и β-коэффициенты для каждого фактора;
3. Коэффициенты частной и множественной корреляции;
4. Значимость множественного уравнения регрессии в целом с помощью общего F-критерия Фишера;
5. Значимость множественных коэффициентов регрессии с использованием критериев Фишера и Стьюдента;
6. Доверительные интервалы множественных коэффициентов регрессии при уровне доверительной вероятности 0,95;

Дать оценку полученным результатам, которые оформить в виде кратких выводов.

По всем вариантам в качестве зависимой переменной (y_1) взять производственную себестоимость производства молока, руб.

Задача 3.9. По данным 20 организаций об удое молока и возрасте коровы (приложение Б) определить: форму связи между признаками; параметры уравнения регрессии; коэффициенты корреляции, детерминации и эластичности. Сделать вывод.

Задача 3.10. По данным 20 организаций об удое молока и содержании в 1 корм. ед. переваримого протеина (приложение Б) определить: форму связи между признаками; параметры уравнения регрессии; тесноту связи между признаками. Сделать вывод.

Задача 3.11. По данным 20 организаций об удое молока и продолжительности сервис-периода (приложение Б) определить: форму связи между признаками; параметры уравнения регрессии; тесноту связи между признаками. Сделать вывод.

Задача 3.12. По данным 20 организаций (приложение Б) с помощью коэффициентов корреляции рангов Спирмена и Кендалла измерить тесноту связи между:

1) возрастом и удоем молока от одной коровы;

2) продолжительностью сервис-периода и удоем молока на одну корову;

3) содержанием в 1 корм. ед. переваримого протеина и удоем молока на одну корову.

Задача 3.13. По данным таблицы 3.12 с помощью коэффициентов ассоциации и контингенции оценить наличие связи между молочной продуктивностью коров и уровнем кормления.

Таблица 3.12 – Распределение поголовья коров по молочной продуктивности и уровню кормления

Уровень кормления	Удой		Всего
	высокий	низкий	
Хороший	150	32	182
Плохой	20	60	80
Итого	170	92	262

Задача 3.14. Установить имеется ли связь между полом помесных кур и пигментацией их ног. В опыте получено 50 петушков и 50 курочек. Из них 5 петушков имели черную окраску ног и 45 – белую, 49 курочек имели пигментированные ноги и 1 курочка оказалась с белыми ногами. С помощью коэффициентов взаимной сопряженности Пирсона и Чупрова оценить наличие связи между полом помесных кур и пигментацией их ног.

Вопросы для самоподготовки

1. Виды и формы связей между признаками.

2. Этапы регрессионного анализа.

3. Что характеризует коэффициент регрессии и способы его расчета при линейной связи.

4. Однофакторный корреляционно-регрессионный анализ.

5. Коэффициенты корреляции, детерминации и эластичности, их интерпретация.

6. Множественная корреляция.

7. Коэффициенты множественной корреляции, детерминации и эластичности, их интерпретация.

8. Ранговая корреляция.

9. Исследование тесноты связи между качественными признаками.

4 Временные ряды

Временной ряд – это совокупность числовых значений изучаемого показателя в последовательные моменты или периоды времени. Он состоит из значений или уровней временного ряда (Y) и периодов или моментов времени (t). Первый член временного ряда называют начальным (y_1), а последний конечным (y_n).

Уровни временного ряда формируются под воздействием большого числа факторов, которые условно можно подразделить на три группы:

– факторы, формирующие тенденцию изменения уровней временного ряда – трендовая компонента (T);

– факторы, формирующие циклические или сезонные колебания уровней ряда – циклическая компонента (S);

– случайные факторы – случайная компонента (ε).

Модель, в которой временной ряд представлен как сумма перечисленных выше компонент, называется аддитивной моделью временного ряда и имеет вид:

$$(Y = T + S + \varepsilon).$$

Если временной ряд представлен как произведение компонент, то она называется мультипликативной моделью временного ряда:

$$(Y = T \cdot S \cdot \varepsilon).$$

Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на один или несколько периодов или моментов времени, называемого коэффициентом автокорреляции.

Коэффициент автокорреляции уровней ряда первого порядка, смещенных на одну единицу времени, определяется по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}} \qquad (4.1)$$

где $\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}$; $\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}$;

Коэффициент автокорреляции уровней ряда второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}, \qquad (4.2)$$

где $\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}$; $\bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}$.

Аналогично можно определить коэффициенты автокорреляции более высоких порядков. Число периодов или моментов времени, по которым рассчитывается коэффициент автокорреляции, называют лагом.

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют аналитическим выравниванием временного ряда. Тенденция во времени может принимать разные формы, для ее формализации используют следующие функции:

– линейная: $y_t = a + bt$;
– степенная: $y_t = at^b$;
– гипербола: $y_t = a + b/t$;
– экспонента: $y_t = e^{a=bt}$;
– показательная: $y_t = ab^t$;
– полином k– ого порядка: $y_t = a + b_1t + b_2t^2 + \ldots + b_k t^k$;
– логическая: $y_t = \frac{1}{1+be^{-ct}}$;
– Гомперца: $log_c f(x) = a - bc^t$, где 0<с<1.

Параметры каждой из перечисленных выше функций определяются методом наименьших квадратов, используя в качестве независимой переменной время (t), а в качестве зависимой переменной – фактические уровни временного ряда (y_t). Для нелинейных трендов предварительно проводят стандартную процедуру линеаризации.

При выборе конкретной функции предпочтение отдается той, которая имеет меньшую сумму квадратов отклонений фактических уровней временного ряда от теоретических, найденных по уравнениям тренда.

Критерием отбора наилучшей формы тренда является значение скорректированного коэффициента детерминации: чем выше его значение, тем лучше форма тренда отражает тенденцию изменения уровней ряда.

Пример 4.1. Имеются данные о продуктивности коров в сельскохозяйственных организациях Краснодарского края, кг/гол.

Таблица 4.2 – Продуктивность коров, кг/гол.

Год	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Продуктивность коров, кг/гол.	3909	4534	4925	5034	5412	5427	5505	5867	6014	6379

Требуется:

1. Построить график динамики продуктивности коров.

2. Рассчитать коэффициент автокорреляции первого порядка.

3. Обосновать выбор типа уравнения тренда и рассчитать его параметры.

4. Дать интерпретацию параметров тренда и сделать выводы по результатам решения.

Решение.

1. Рассмотрим систему координат, где Y_t – продуктивности коров, ц/гол.; t – порядковый номер года и нанесем в ней данные таблицы 4.2 на график (рисунок 4.1).

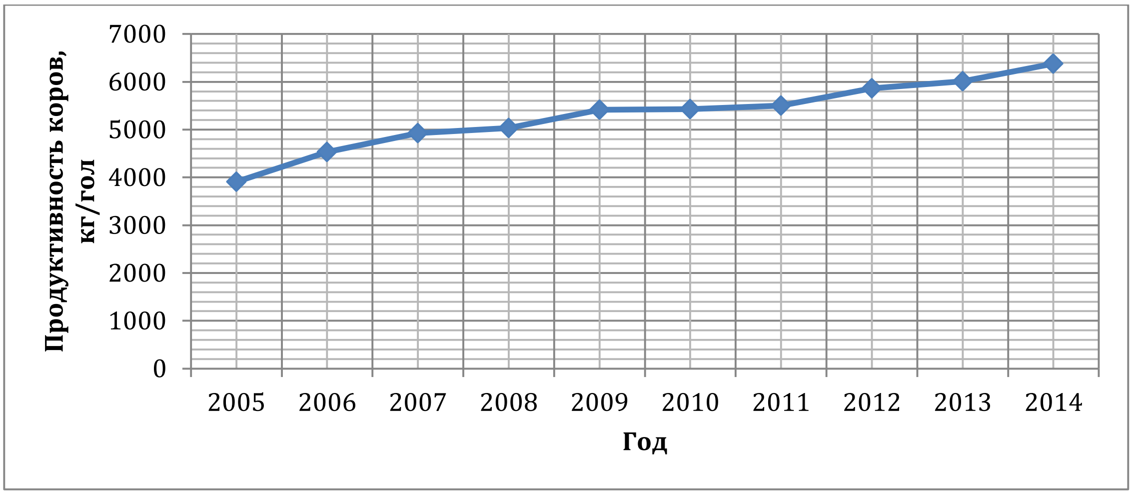

Рисунок 4.1 – Динамика продуктивности коров

2. Определим коэффициент автокорреляции первого порядка, характеризующего степень тесноты связи между последовательными уровнями временного ряда продуктивности коров, сдвинутыми на один год. Заполним вспомогательную таблицу 4.3.

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{53006 - 3909}{9} = 5455{,}22;$$

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{46627}{9} = 5180{,}78;$$

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}} = \frac{2921005{,}4}{\sqrt{2647515{,}6 \cdot 3507075{,}6}} = 0{,}959.$$

Таблица 4.3 – Вспомогательная таблица для расчета коэффициента автокорреляции

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \overline{y_1}) \cdot (y_{t-1} - \overline{y_2})$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	3909	–	–	–	–	–	–
2	4534	3909	–921,22	–1271,78	1171590,0	848650,4	1617418,7
3	4925	4534	–530,22	–646,78	342936,0	281135,6	418321,5
4	5034	4925	–421,22	–255,78	107739,3	177428,2	65422,3
5	5412	5034	–43,22	–146,78	6344,1	1868,2	21543,7
6	5427	5412	–28,22	231,22	–6525,6	796,5	53463,7
7	5505	5427	49,78	246,22	12256,4	2477,8	60625,4
8	5867	5505	411,78	324,22	133507,5	169560,9	105120,0
9	6014	5867	558,78	686,22	383445,7	312232,6	470900,9
10	6379	6014	923,78	833,22	769712,2	853365,4	694259,3
Сумма	53006	46627	0,00	0,00	2921005,4	2647515,6	3507075,6

3. Полученное значение коэффициента автокорреляции и графическое изображение временного ряда позволяют сделать вывод о том, что ряд продуктивности коров содержит тенденцию, близкую к линейной. Поэтому для моделирования его тенденции используем линейную функцию:

$$y = a + bt.$$

Для расчета параметров линейного тренда a и b используем метод наименьших квадратов, для чего составим и решим следующую систему:

$$\begin{cases} \sum y = na + b\sum t \\ \sum yt = a\sum t + b\sum t^2 \end{cases}$$

Для составления и решения системы уравнений заполним таблицу 4.4.

Таблица 4.4 – Вспомогательная таблица для расчета параметров тренда

№ п/п	*y*	*t*	*Yt*	t^2	$\hat{y}_t$	$(y-\hat{y}_t)^2$
1	3909	1	3909	1	4245	112896
2	4534	2	9068	4	4480	2916
3	4925	3	14775	9	4715	44100
4	5034	4	20136	16	4950	7056
5	5412	5	27060	25	5185	51529
6	5427	6	32562	36	5420	49
7	5505	7	38535	49	5655	22500
8	5867	8	46936	64	5890	529
9	6014	9	54126	81	6125	12321
10	6379	10	63790	100	6360	361
Сумма	53006	55	310897	385	53025	254257
Среднее значение	5300,6	5,5	31089,7	38,5	-	-

Воспользуемся формулами, вытекающими из системы:

$$b = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\overline{t^2} - (\bar{t})^2} = \frac{31089{,}7 - 5300{,}6 \cdot 5{,}5}{38{,}5 - 5{,}5^2} = 235$$

$$a = \bar{y} - b\bar{t} = 5300{,}6 - 235 \cdot 5{,}5 = 4010.$$

Уравнение прямой примет вид:

$$Y = 4010 + 235t.$$

Таким образом, в среднем ежегодно за 2005–2014 гг. продуктивность коров увеличивалась на 235 ц/гол.

Рассчитаем прогнозное значение средней продуктивности коров на 2017 г. путем подстановки в уравнение линейного тренда значения $t_{пр}$=13:

$$\hat{Y}_{пр} = 4010 + 235 \cdot 13 \ \rightarrow \hat{Y}_{пр} = 7065 \text{ ц/гол.}$$

Однако точечный прогноз нереален и он дополняется расчетом интервальной оценки $\hat{y}^*_{пр}$ с учетом 95 %-й доверительной вероятности:

$$\hat{Y}_{пр} - t_\alpha \cdot m_{\hat{y}_t} \le y^*_{пр} \le \hat{Y}_{пр} + t_\alpha \cdot m_{\hat{y}_t},$$

где t_α – критическое значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $df = n - 2$;

$$m_{\hat{y}_t} = \sqrt{S^2_{ост}\left(\frac{1}{n} + \frac{(t_{пр}-\bar{t})^2}{\sum(t-\bar{t})^2}\right)}$$ – стандартная ошибка прогноза.

$$S^2 = \frac{\sum(y-\hat{y}_t)^2}{n-2} = \frac{254257}{8} = 31782{,}13,$$

$$m_{\hat{y}_t=}\sqrt{31782{,}13 \cdot \left(\frac{1}{10} + \frac{(13-5.5)^2}{82{,}5}\right)} = 157{,}7$$

при $t_{\alpha=0.05;df=8} = 2{,}3.$

$$7065 - 2{,}3 \cdot 157{,}7 \le y^*_{пр} \le 7065 + 2{,}3 \cdot 157{,}7,$$

$$6702 \le y^*_{пр} \le 7428.$$

Следовательно, с доверительной вероятностью 0,95 можно утверждать, что продуктивность коров в 2017 г. будет находиться в интервале от 6702 до 7428 ц/гол.

Задача 4.1. По имеющимся данным о поголовье крупного рогатого скота в сельскохозяйственных предприятиях: а) построить уравнение линейного тренда; б) дать интерпретацию параметров линейного тренда; в) определить коэффициент детерминации для линейного тренда.

Год, t	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Поголовье скота, тысяч голов,(у)	733	729	673	582	520	490	478	470	460	440

Задача 4.2. Используя данные по валовому надою молока в сельскохозяйственных организациях Краснодарского края обосно-

вать выбор уравнения тренда, рассчитать параметры выбранного тренда и дать прогноз о количестве крупного рогатого скота в процентах в общей структуре поголовья скота в 2015 г.

Год , *t*	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Валовой надой молока, тыс. т, (у)	1004	977	908	891	885	869	853	905	870	851

Задача 4.3 По данным по производству мяса (в убойном весе) в сельскохозяйственных хозяйствах всех категорий Краснодарского края построить экспоненциальный тренд, дать интерпретацию его параметров.

Год , *t*	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Мясо, тыс. т (*y*)	287	305	301	297	360	368	378	375	394	429

Задача 4.4 По данным по производству меда в сельскохозяйственных предприятиях Краснодарского края рассчитать коэффициента автокорреляции 1-го и 2-го порядков, обосновать выбор уравнения тренда, дать прогноз производства меда на 2016 г.

Год , *t*	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Мед, тонн	3474	3505	3508	3605	3570	3475	2999	3059	2853	2586

Задача 4.5. По данным приложения В по одному виду заболевания за последние пять лет построить график динамики заболеваемости, рассчитать коэффициент автокорреляции первого порядка, обосновать выбор типа уравнения тренда и рассчитать его параметры, дать интерпретацию параметров тренда и сделать выводы по результатам решения.

Задача 4.6. Выявить общую тенденцию заболеваемости животных по данным приложения В, используя приемы укрупнения периодов, трехлетней скользящей средней и аналитического выравнивания. Изобразить на графике фактические и выравненные (теоретические) уровни. Сделать вывод по результатам расчетов.

Задача 4.7. По данным задачи 4.6 выполнить прогноз уровня заболеваемости на предстоящие два года.

Вопросы для самоподготовки

1. Ряды динамики, их элементы и правила построения.
2. Виды рядов динамики.
3. Статистические показатели ряда динамики и порядок их расчета.
4. Средние показатели в рядах динамики и порядок их расчета.
5. Приемы изучения тенденции в рядах динамики.
6. Выравнивание ряда способом наименьших квадратов.
7. Автокорреляция уровней ряда.

Список рекомендуемой литературы

1. Аскеров, П. Ф. Общая и прикладная статистика: учебник / П. Ф. Аскеров, Р. Н. Пахунова, А. В. Пахунов. – ИНФРА-М, 2014. – 271 с.

2. Дегтярева И.Н. Статистика. Общая теория [Электронный ресурс]: учебно-практическое пособие/ И.Н. Дегтярева— Электрон. текстовые данные.— Саратов: Вузовское образование, 2015.— 183 с.— Режим доступа: http://www.iprbookshop.ru/37224.html.— ЭБС «IPRbooks» по паролю

3. Лисицин, Д. В. Устойчивые методы оценивания параметров статистических моделей [Электронный ресурс] : учеб. пособие / Д. В. Лисицин – Электрон. текстовые данные. – Новосибирск : Новосибирский государственный технический университет, 2013. – 76 с. – Режим доступа: http://www.iprbookshop.ru/45452. – ЭБС «IPRbooks», по паролю

4. Ляховецкий, А. М. Статистика : учеб. пособие / Ляховецкий А.М., Кремянская Е.В., Климова Н.В. ; под ред. В.И. Нечаева. - М. : КноРус, 2016. - 362 с.

5. Рафикова Н. Т. Основы статистики [Электронный ресурс] : учеб. пособие / Н. Т. Рафикова. – Электрон. текстовые данные. – М.: Финансы и статистика, 2014. – 352 с. – Режим доступа: http://www.iprbookshop.ru/18824. – ЭБС «IPRbooks», по паролю.

6. Теория статистики [Электронный ресурс] : учебник / Р. А. Шмойлова [и др.]. – Электрон. текстовые данные. – М. : Финансы и статистика, 2014. – 656 с. – Режим доступа: http://www.iprbookshop.ru/18846. – ЭБС «IPRbooks», по паролю.

7. Улитина, Е.В. Статистика [Электронный ресурс]: учебное пособие/ Улитина Е.В., Леднева О.В., Жирнова О.Л.— Электрон. текстовые данные.— М.: Московский финансово-промышленный

университет «Синергия», 2013.— 320 с.— Режим доступа: http://www.iprbookshop.ru/17045.— ЭБС «IPRbooks», по паролю

8. Федин Ф. О. Анализ данных. Ч. 1. Подготовка данных к анализу [Электронный ресурс] : учеб. пособие / Ф. О. Федин, Ф. Ф. Федин. – Электрон. текстовые данные. – М. : Московский городской педагогический университет, 2012. – 204 с. – Режим доступа: http://www.iprbookshop.ru/26444. – ЭБС «IPRbooks», по паролю.

Приложение А

Таблица А 1 – Данные по производству молока в сельскохозяйственных предприятиях северной зоны Краснодарского края, 2011 г.

№ п/п	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6
1	1242	1260	5488	3,18	700	118,0	646,7	74,5
2	1543	1560	3800	5,22	257	68,5	468,7	49,9
3	1421	1422	5453	1,69	500	162,3	610,6	78,6
4	1154	1214	6231	1,43	450	276,7	613,5	79,5
5	1366	1366	4616	3,61	1440	64,8	705,2	63,0
6	1746	1741	4939	2,62	575	119,1	549,7	64,2
7	1118	1166	5040	2,83	736	141,9	386,9	69,4
8	1381	1433	5201	2,49	1150	110,0	730,5	61,0
9	1424	1474	4636	3,11	500	156,9	717,0	74,0
10	1375	1413	5179	2,82	787	168,4	810,2	74,2
11	1520	1540	3732	3,17	600	113,0	661,0	64,1
12	1419	1419	5633	3,64	1340	124,4	644,3	81,1
13	1117	1194	3632	1,65	300	144,2	440,0	84,5
14	1207	1310	6711	1,47	2500	155,6	628,5	60,1
15	1284	1444	5602	1,52	2000	251,5	402,4	34,8
16	1458	1462	5741	2,15	1750	128,6	704,1	67,4
17	855	875	5527	3,88	2000	70,6	351,5	37,3
18	1120	1121	7407	1,88	1900	108,0	590,9	81,0
19	1193	1326	7561	1,80	823	106,0	650,5	90,9
20	1263	1264	4208	3,18	836	98,4	525,3	66,2
21	1177	1249	6520	2,06	358	157,7	619,9	72,5
22	1175	1212	5195	1,73	1200	98,2	583,0	84,9
23	1465	1510	4367	3,20	229	124,6	612,0	88,2
24	1359	1386	4539	2,51	800	206,1	652,0	77,8
25	1387	1396	4071	5,11	620	39,2	626,3	93,7
26	1058	1078	3714	2,94	403	115,4	379,0	80,9
27	1431	1436	9655	1,06	1233	162,4	598,0	90,8
28	2502	2582	2718	7,53	918	62,9	1475,4	49,5
29	1305	1305	6902	2,32	905	110,6	621,3	83,2
30	1291	1295	5904	2,36	640	104,8	669,0	79,0
31	1235	1318	8003	1,02	622	124,4	509,1	95,2
32	1624	1628	3431	3,63	632	58,8	610,6	28,4
33	1324	1391	4914	2,46	2992	159,7	461,9	74,2
34	1186	1186	7434	1,28	746	207,9	501,1	85,9
35	1274	1284	4455	1,40	595	123,9	568,0	81,3
36	1136	1136	5145	3,54	500	58,8	418,1	72,8
37	1773	1773	3883	4,11	670	89,3	774,3	61,0
37	1178	1165	7236	1,57	2108	306,9	436,9	39,8
39	1088	1001	7329	2,30	600	169,1	390,2	45,7
40	1493	1496	4410	2,56	850	110,3	794,7	69,2

Обозначения: y_1 – производственная себестоимость производства молока, руб.
y_2 – коммерческая себестоимость производства молока, руб.
x_1 – годовой надой молока на среднегодовую корову, кг;
x_2 – прямые затраты труда на 1 ц, чел.-ч;
x_3 – среднегодовое поголовье коров на предприятии, гол.;
x_4 – затраты по оплате труда на 1 ц, руб.;
x_5 – затраты на корма на 1 ц молока, руб.;
x_6 – доля молока в выручке от реализации продукции животноводства, %.

Приложение Б

Таблица Б 1 – Показатели молочного скотоводства в сельскохозяйственных организациях

Номер предприятия	Продолжительность сервис-периода, дн	Возраст коровы, лет	Содержание в 1 корм. ед. переваримого протеина, г	Поголовье, гол.	Удой молока от одной коровы, ц/гол.
1	68	5,4	156	3115	52,0
2	80	7,2	156	4010	63,0
3	55	4,8	158	2500	50,0
4	45	3,0	84	4800	41,0
5	87	7,3	149	4085	67,1
6	88	7,0	145	3510	68,2
7	90	6,3	280	4570	69,0
8	78	5,7	134	3715	56,1
9	65	4,9	163	2935	52,0
10	70	5,1	115	3025	55,3
11	64	5,0	97	4015	48,4
12	61	5,2	157	5014	55,5
13	51	4,3	81	2601	44,0
14	63	4,7	103	3021	48,0
15	66	5,1	115	3075	50,5
16	88	7,4	300	3105	68,0
17	48	6,2	164	3401	58,1
18	80	7,5	280	4010	66,0
19	94	7,0	320	3412	69,5
20	76	6,9	250	4210	64,0
21	53	5,5	97	1650	46,5
22	64	4,6	97	3850	48,9
23	80	7,0	140	3005	66,0
24	86	6,0	260	3170	61,0
25	50	4,3	115	2075	44,0
26	57	5,1	130	3510	52,5
27	81	6,0	290	3110	62,4
28	92	6,2	280	3940	66,0
29	75	8,0	255	2201	66,4
30	58	3,5	75	1217	43,5
31	66	4,0	160	2515	45,0
32	55	4,5	102	1941	42,6
33	58	5,0	108	3805	49,5

Продолжение таблицы Б 1

Номер предприятия	Продолжительность сервис-периода, дн	Возраст коровы, лет	Содержание в 1 корм. ед. переваримого протеина, г	Поголовье, гол.	Удой молока от одной коровы, ц/гол.
34	75	5,5	146	3640	56,5
35	60	4,0	188	3200	42,0
36	45	4,5	105	3810	42,0
37	80	5,5	260	4015	57,1
38	89	6,0	230	3104	55,0
39	90	7,4	275	4575	63,0
40	78	6,2	134	3620	52,0
41	65	5,1	172	2940	43,4
42	68	5,4	165	3101	45,2
43	67	4,9	101	4010	49,9
44	61	4,6	157	5010	45,8
45	52	3,0	101	2540	40,0
46	63	3,5	102	3100	49,0
47	65	4,0	115	2957	40,5
48	86	6,0	300	3210	66,5
49	48	3,2	156	3201	46,5
50	80	6,4	275	3975	62,7
51	94	7,1	320	3415	68,5
52	75	7,0	245	3450	62,1
53	50	3,8	159	1680	45,2
54	64	4,0	98	3420	48,0
55	80	5,0	145	2975	57,4
56	85	6,5	260	3150	60,1
57	69	3,9	110	2089	45,0
58	75	3,8	130	3640	50,5
59	79	6,6	290	3140	63,0
60	92	7,0	285	3450	68,0
61	74	7,5	245	2117	60,7
62	65	6,2	160	2417	56,4
63	52	4,5	122	2515	44,1
64	68	3,5	156	3115	42,0
65	80	5,0	156	4010	53,0
66	55	4,5	158	2500	40,0
67	45	4,6	84	4800	41,0
68	87	6,1	149	4085	60,1

Приложение В

Таблица В 1 – Число зарегистрированных случаев болезни животных в хозяйствах региона, гол.

Вариант	Болезнь	Год								
		2007	2008	2009	2010	2011	2012	2013	2014	2015
Крупный рогатый скот										
1	Диспепсия телят	442	481	427	451	553	457	497	551	398
2	Псевдомоноз телят	225	275	238	225	342	268	302	319	263
3	Кетоз	465	420	435	475	514	469	492	538	371
4	Атония преджелудка	234	242	258	187	369	270	250	338	274
5	Ацидоз	233	295	222	264	447	321	385	447	312
6	Родильный парез	481	441	402	218	495	338	338	477	419
7	Трихофития	263	266	252	210	337	233	247	305	252
8	Папиллома	99	134	144	125	212	64	166	240	168
9	Лейкоз	43	64	66	45	102	49	98	71	70
10	Мастит	397	444	471	483	507	603	621	610	643
11	Колибациллез	233	197	229	144	329	231	237	279	411
12	Паратиф	236	198	233	146	339	236	240	281	219
13	Тимпания	396	328	359	262	4386	381	361	438	423
14	Сальмонеллёз	180	193	187	164	221	202	194	219	214
15	Межпальцевая флегмона	182	209	207	189	233	209	208	233	232
16	Флегмона венчика	180	126	128	91	158	180	151	185	180
17	Пупочный сепсис	166	147	169	162	191	178	195	201	161
18	Дифтерия	928	884	896	789	968	939	890	964	986
19	Лейшманиоз	74	85	93	79	102	106	98	111	106
20	Гастроэнтерит	432	632	662	499	687	865	666	637	848
21	Рахит	325	260	247	196	251	256	284	239	227
22	Хламидиоз	125	126	141	122	153	138	136	141	1239
Свиньи										
23	Стоматит	732	640	809	655	966	916	785	669	815
24	Авитаминоз	193	155	171	126	189	166	140	193	155
25	Закупорка пищевода	192	276	349	219	306	250	238	304	259
26	Колибациллез	95	86	101	79	113	87	93	91	69
27	Сальмонеллез	264	236	269	165	307	326	462	437	364
28	Диспепсия поросят	115	93	121	69	133	85	95	97	104
29	Гастроэнтерит	457	578	491	555	704	699	605	744	873
30	Катар ЖКТ	561	686	486	807	740	845	795	1134	756

Приложение Г

Таблица Г1 – Критические точки t-распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,31	636,62
2	2,92	4,30	6,97	9,93	22,33	31,60
3	2,35	3,18	4,54	5,84	10,22	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,90	2,37	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,06	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,15	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,02
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,70	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,47	2,76	3,41	3,67
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,16	3,37
∞	1,65	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Приложение Д

Таблица Д 1 – Критические точки χ^2 – распределения Пирсона

ν \ α	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19.679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45.315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

Приложение Е

Таблица Ж 1 – Таблица вероятностей для t-критерия Стьюдента

n	Доверительная вероятность			
	0,90	0,95	0,99	0,999
2	6,3	12,7	63,7	636,6
3	2,9	4,3	9,9	31,6
4	2,4	3,2	5,8	12,9
5	2,1	2,8	4,6	8,6
6	2,0	2,6	4,0	6,9
7	1,9	2,4	3,7	6,0
8	1,9	2,4	3,5	5,4
9	1,9	2,3	3,4	5,0
10	1,8	2,3	3,2	4,8
11	1,8	2,2	3,2	4,6
12	1,8	2,2	3,1	4,4
13	1,8	2,2	3,1	4,3
14	1,8	2,2	3,0	4,2
15	1,8	2,1	3,0	4,1
16	1,8	2,1	2,9	4,1
17	1,7	2,1	2,9	4,0
18	1,7	2,1	2,9	4,0
19	1,7	2,1	2,9	3,9
20	1,7	2,1	2,9	3,9
21	1,7	2,1	2,8	3,8
22	1,7	2,1	2,8	3,8
23	1,7	2,1	2,8	3,8
24	1,7	2,1	2,8	3,8
25	1,7	2,1	2,8	3,7
26	1,7	2,1	2,8	3,7
27	1,7	2,1	2,8	3,7
28	1,7	2,1	2,8	3,7
29	1,7	2,0	2,8	3,7
30	1,7	2,0	2,8	3,7

Оглавление

У ч е б н о е   и з д а н и е

Кацко Игорь Александрович,
Горпинченко Ксения Николаевна,
Захарова Юлия Николаевна и др.

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Практикум

В авторской редакции

Подписано в печать __.__.2017. Формат 60 × 84 $^{1}/_{16}$.
Усл. печ. л. – 5,2. Уч.-изд. л. – 4,0.
Тираж 50. Заказ №____

Типография Кубанского государственного аграрного университета
имени И. Т. Трубилина
350044, г. Краснодар, ул. Калинина, 13