

Воронежский институт МВД России

О. Ю. Данилова
В. А. Родин
С. В. Синегубов

ПРАВОВАЯ СТАТИСТИКА

Учебное пособие

Воронеж
2023

ББК 67.54
УДК 311 : 519.4

Рецензенты:

Кобзистый С. Ю. — доцент кафедры технических комплексов охраны и связи ФКОУ ВО Воронежский институт ФСИН России, кандидат технических наук, доцент, подполковник внутренней службы;

Самороковский А. Ф. — доцент кафедры гуманитарных и социально-экономических дисциплин Центрального филиала Российского государственного университета правосудия, кандидат технических наук, доцент.

Данилова О. Ю.

Правовая статистика : учебное пособие [Электронный ресурс] / О. Ю. Данилова, В. А. Родин, С. В. Синегубов. — Электр. дан. и прогр. — Воронеж : Воронежский институт МВД России, 2023. — 1 электр. опт. диск (CD-ROM) : 12 см. — Систем. требования : процессор Intel с частотой не менее 1,3 ГГц ; ОЗУ 512 Мб ; операц. система семейства Windows ; CD-ROM дисковод.

Рассматриваются классические модели линейной регрессии, анализ временных рядов. Приведены необходимые теоретические сведения и подробно рассматривается решение задач прикладной статистики с использованием пакета ASTRA LINUX: LibreOffice: Calc. Приведены задачи для самостоятельного решения.

Учебное пособие предназначено для слушателей факультета заочного обучения по специальности 40.05.02 Правоохранительная деятельность и по направлениям подготовки 40.03.02 Обеспечение законности и правопорядка, 40.03.01 Юриспруденция.

Пособие может быть полезно сотрудникам органов внутренних дел, деятельность которых связана с анализом и обработкой статистической информации.

ISBN 978-5-88591-966-1
© Воронежский институт МВД России, 2023

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
1. ПОНЯТИЕ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ	
1.1. Пример модели правовой статистики	5
1.2. Понятие и основные типы моделей	6
1.3. Расчет сводных характеристик выборок и их взаимосвязей	9
2. УРАВНЕНИЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ	
2.1. Постановка задачи построения модели множественной регрессии	11
2.2. Оценка качества уравнения регрессии	12
2.3. Выбор вида уравнения парной регрессии	14
2.4. Оценка параметров уравнения линейной парной регрессии	17
3. МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ	
3.1. Структура временного ряда	20
3.2. Выявление компонент временного ряда	20
4. ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ	
4.1. Пример 1	23
4.2. Пример 2	31
4.3. Пример 3	42
5. КОНТРОЛЬНЫЕ ВОПРОСЫ	71
6. ВАРИАНТЫ ЗАДАНИЙ	
5.1. Задача 1	73
5.2. Задача 2	77
5.3. Задача 3	83
ЛИТЕРАТУРА	87

ВВЕДЕНИЕ

Сегодня деятельность в любой области права требует от специалиста применения современных методов работы, использующих информационные технологии. Значительная часть этих методов связана с использованием статистических моделей. Изучение этих методов осуществляется в рамках курса правовой статистики.

Правовая статистика — прикладная юридическая наука, изучающая количественные показатели, характеризующие динамику правовых процессов.

Важное значение имеет правовая статистика в криминологических исследованиях, поэтому главными ее разделами являются уголовная статистика и судебная статистика.

Специфической особенностью деятельности юриста является работа в условиях недостатка информации и неполноты исходных данных. Анализ такой информации требует специальных методов, которые составляют один из аспектов правовой статистики.

Данный курс ставит своей задачей сообщить слушателю сведения по правовой статистике, необходимые для успешного освоения дисциплины, а также развить навыки логического мышления. Настоящее пособие имеет целью помочь слушателю заочной формы обучения в самостоятельной работе над учебным материалом.

В пособии подробно рассматривается решение задач прикладной статистики с использованием пакета ASTRA LINUX: LibreOffice: Calc.

Выполнение домашней контрольной работы является составной частью самостоятельной работы слушателя-заочника и служит целям закрепления теоретических знаний и приобретения практических навыков использования статистических методов в практической деятельности.

В разделах 1-3 даётся краткое описание теоретического материала, в разделе 4 описаны примеры составления и использования моделей правовой статистики, а задания для самостоятельного выполнения приведены в разделе 5.

Контрольная работа содержит три задания для самостоятельного выполнения, для каждого разработано 10 вариантов. Номер варианта – последняя цифра в порядковом номере списка учебной группы в журнале (преподаватель вправе изменить номер варианта). Работа оценивается по системе «зачтено» или «не зачтено». Если контрольная работа не зачтена, слушатель должен внимательно изучить рецензию на контрольную работу, исправить все недостатки, указанные преподавателем, и прислать на повторную проверку исправленные задания совместно с первоначальной работой.

Все примеры и задания носят академический характер, никаких социальных выводов по данным примерам и задачам делать не стоит.

Авторы выражают благодарность доценту кафедры математики и моделирования систем А. Н. Копылову за ценные замечания и предложения, высказанные при написании данного издания.

1. ПОНЯТИЕ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

1.1. Пример модели правовой статистики

Рассмотрим задачу экспертной оценки прогнозируемого количества лиц, которые могут совершить преступления в некотором районе. Эта задача, не претендуя на общность, позволяет выявить основные особенности математических моделей.

Чтобы произвести указанную оценку, необходимо найти выборку подходящих наблюдений по другим районам.

Пусть i - номер наблюдения. Обозначим

x_1^i – количество жителей в районе;

x_2^i – количество ранее судимых;

x_3^i – количество стоящих на учете подростков;

y^i – количество лиц, совершивших преступления.

Вся выборка имеет вид набора четверок:

$$\begin{aligned} & y^1, x_1^1, x_2^1, x_3^1, \\ & y^2, x_1^2, x_2^2, x_3^2, \\ & \dots, \\ & y^n, x_1^n, x_2^n, x_3^n, \end{aligned} \quad (1.1)$$

где n – общее число наблюдений.

Проведя анализ наблюдений по другим районам, эксперт находит оценку показателя для интересующего его района. Это означает, что он неявно построил некоторую модель в виде функциональной зависимости

$$\hat{y} = f(x_1, x_2, x_3), \quad (1.2)$$

где $\hat{y}$ – искомая оценка количества лиц, совершивших преступления в зависимости от заданных значений параметров x_1, x_2, x_3 .

Подставив значения параметров x_1^0, x_2^0, x_3^0 для своего района, может получить $\hat{y}^0 = f(x_1^0, x_2^0, x_3^0)$ – экспертную оценку предполагаемого количества лиц, которые совершат преступление.

При этом эксперт старается сделать так, чтобы построенная модель как можно лучше соответствовала наблюдениям, т.е. чтобы

$$\hat{y}^i = f(x_1^i, x_2^i, x_3^i) \approx y^i.$$

Здесь $\hat{y}^i$ – теоретическое, т.е. ожидаемое значение оценки в соответствии с моделью (1.2), а y^i – фактическое значение оценки в этом наблюдении.

Можно утверждать, что теоретическое и фактическое значения оценки практически никогда не будут совпадать потому, что выборка (1.1), как правило, является противоречивой. Например, она может содержать одинаковые значения параметров x_1, x_2, x_3 и при этом различные значения результата y . Если использовать язык математики, это означает, что y – случайная величина. В тоже время $\hat{y}$ –

величина детерминированная, однозначно определяемая значениями x_1, x_2, x_3 . Поэтому можно записать

$$y = \hat{y} + \varepsilon = f(x_1, x_2, x_3) + \varepsilon,$$

где ε — случайная составляющая, в которой

- 1) суммировались другие, не учтенные экспертом факторы (например, количество сотрудников полиции, их техническая оснащенность и т.п.),
- 2) сказалась неточность модели (1.2).

Чем меньше значение $|\varepsilon|$, тем более точной является модель (1.2).

На данном примере можно представить основные особенности моделирования:

1) случайный характер значений отдельных показателей определяет выбор теории вероятностей и математической статистики как основного математического аппарата для моделирования;

2) основными задачами, решаемыми в процессе моделирования, являются:

- выбор множества независимых переменных, обычно называемых факторами $X = \{x_1, x_2, \dots, x_m\}$, по которым осуществляется определение интересующего нас множества зависимых переменных, обычно называемых результатами $Y = \{y_1, y_2, \dots, y_p\}$;

- выбор вида модели, т. е. функциональной зависимости $Y = f(X)$ так, чтобы случайная погрешность $|\varepsilon|$ была наименьшей.

1.2. Понятие и основные типы моделей

Правовая статистика позволяет определять количественное выражение взаимосвязей явлений и процессов, выраженные через некоторые показатели, связанные с правовыми нарушениями. Эти показатели имеют следующую особенность – одним и тем же значениям одних показателей могут соответствовать различные значения других показателей.

Например,

1) в районах с приблизительно одинаковым числом жителей число совершаемых преступлений может существенно различаться;

2) число преступлений в одном районе в различные периоды времени различно.

Поэтому основным инструментом исследования таких показателей являются математические модели, разработанные на основе методов математической статистики, в основе которой лежит выборочный метод.

Суть метода состоит в следующем.

Под **генеральной совокупностью** понимают множество объектов, подлежащих статистическому исследованию. Как правило, исследовать всю эту совокупность либо невозможно, либо затруднительно. Поэтому из этой совокупности выбирают небольшое количество — **выборку**, после изучения которой делают вывод обо всей совокупности. Выборка должна быть сформирована так, чтобы правильно отражать представление о генеральной совокупности, т.е. быть **репрезентативной**.

Таким образом, исходными данными для построения моделей правовой статистики являются **выборки наблюдений**, т. е. наборы значений некоторых показателей (переменных), например:

- 1) количество жителей района,
- 2) количество ранее судимых,

3) количество лиц, совершивших преступления.

Выборки наблюдений подразделяют на два типа:

пространственные выборки, когда наблюдения относятся к различным объектам, определенным в один и тот же момент времени;

временные выборки, когда наблюдения относятся к одному объекту, но в различные моменты времени.

Задачей статистического моделирования является прогноз будущих значений показателей на основе исследования взаимоотношений между показателями и выборками наблюдений их значений.

Модели, построенные по временным выборкам, сложнее моделей, построенных по пространственным. Это связано с тем, что временные выборки могут отражать тенденции изменения значений показателей и циклические колебания этих значений.

Для того, чтобы прогнозировать значения одних показателей, как правило, необходимо знать значения других. Поэтому показатели разделяют на

независимые (экзогенные), т. е. внешние, задаваемые пользователем, **зависимые (эндогенные)**, т. е. внутренние, определяемые в модели.

Значения показателей подразделяются на:

- лаговые, датируемые предыдущими моментами времени;
- текущие, датируемые настоящим моментом времени;
- ожидаемые, датируемые будущими моментами времени.

Каждая модель разрабатывается с определенной целью. Основными целями статистического моделирования обычно являются:

- анализ исследуемого объекта, процесса или явления, имитация его развития при различных значениях независимых показателей;

- прогноз значений показателей.

Выделяют четыре основных этапа:

- спецификация,
- информационный,
- идентификация,
- верификация.

На этапе **спецификации** осуществляется глубокое качественное изучение моделируемого объекта (процесса, явления). На этой основе определяется состав эндогенных и экзогенных показателей, характер связи между ними. Это позволяет определить тип модели.

Информационный этап предполагает сбор информации, представляющей собой выборки наблюдений.

Идентификация модели заключается в проведении статистического анализа модели и оценки неопределенных параметров модели. Если осуществление этого этапа оказывается невозможным, то либо неправильно осуществлена спецификация, либо на информационном этапе собрано недостаточно данных.

Верификация — это проверка истинности модели, т. е. ее адекватности исследуемому объекту, процессу или явлению. Если осуществление этого этапа оказывается невозможным, то либо неправильно осуществлена спецификация, либо на информационном этапе собрано недостаточно данных.

Если исходные данные представляют собой пространственные выборки, то для нахождения значений зависимых показателей по независимым строят **уравнения регрессии**.

Если исходные данные представляют собой временные данные, то для нахождения будущих значений показателей по текущим и лаговым строят модели **временных рядов**.

Методы построения этих двух типов моделей мы в дальнейшем и будем рассматривать.

1.3. Расчет сводных характеристик выборок и их взаимосвязей

Пусть задана выборка

$$\{x_1, x_2, \dots, x_n\} \quad (1.3)$$

значений показателя X .

Для того, чтобы в сжатой форме охарактеризовать эту выборку, используют её числовые характеристики. Наиболее часто при разработке статистических моделей используют

- выборочную среднюю,
- выборочную дисперсию,
- выборочное среднееквадратическое отклонение.

Выборочная средняя $\bar{X}$ определяется по формуле:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}.$$

Выборочная дисперсия $D(X) = \sigma_X^2$ определяется по формуле:

$$D(X) = \sigma_X^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}.$$

Выборочным среднееквадратическим отклонением σ_X выборки X называется арифметический квадратный корень из ее выборочной дисперсии, т.е. $\sigma_X = \sqrt{D(X)}$.

И дисперсия и среднееквадратическое отклонение характеризуют разброс значений выборки относительно ее среднего значения. Но $D(X)$ имеет размерность квадрата случайной величины, а σ_X — размерность самой случайной величины.

Зависимость между двумя показателями называется вероятностной (стохастической, статистической), если изменение значений одного из них приводит к изменению значений другого. Если же изменение одного показателя не влияет на изменение другого, то эти показатели называются независимыми.

Примером вероятностной зависимости является зависимость между количеством несовершеннолетних, состоящих на учете и количеством совершенных ими преступлений.

Пусть имеются основания считать, что два показателя, заданные выборками X и Y , являются взаимосвязанными. Рассмотрим вопрос об оценке этой взаимосвязи.

Говорят, что эта связь является линейной, если для соответствующих значений показателей выполняется соотношение

$$Y = a + bX + \varepsilon, \quad (1.4)$$

где ε — некоторая остаточная величина.

В этом случае используются ковариация, коэффициент корреляции, коэффициент детерминации.

Ковариация $\text{cov}(X, Y)$ рассчитывается по формуле:

$$\text{cov}(X, Y) = \overline{(X - \bar{X})(Y - \bar{Y})}.$$

Коэффициент корреляции r_{XY} рассчитывается по формуле:

$$r_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Коэффициент детерминации r_{XY}^2 рассчитывается по формуле:

$$r_{XY}^2 = (r_{XY})^2.$$

Перечисленные показатели обладают следующими свойствами:

- 1) $\text{cov}(X, Y) = \overline{XY} - \bar{X}\bar{Y}$;
- 2) $-1 \leq r_{XY} \leq 1$;
- 3) если $\text{cov}(X, Y) = r_{XY}$ близки к 0, то показатели X и Y независимы;
- 4) чем ближе $|r_{XY}|$ и r_{XY}^2 к 1, тем более тесная линейная связь между X и Y ;
- 5) чем ближе $|r_{XY}|$ и r_{XY}^2 к 0, тем линейная связь между X и Y менее выражена.

Зависимость между показателями X и Y может носить не только линейный характер.

2. УРАВНЕНИЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

2.1. Постановка задачи построения модели множественной регрессии

Задачами регрессионного анализа являются установление функциональной зависимости между зависимыми и независимыми показателями и оценка значений зависимых показателей по значениям независимых, т.е. прогноз значений этих показателей. Пример был приведён в разделе 1.

Уравнение множественной регрессии характеризует связь сразу нескольких показателей $Y, X_1, X_2, \dots, X_m$, где

Y — зависимый показатель, в дальнейшем называемой **результатом**,

$X_1, X_2, \dots, X_m$ — независимые показатели, в дальнейшем называемые **факторами**.

Связь между переменными проявляется как некоторая закономерность лишь в среднем на основе выборки совместных наблюдений, т.е. наборов значений зависимого и независимых показателей, относящихся к одному объекту:

$$\begin{aligned} & (y^1, x_1^1, x_2^1, \dots, x_m^1), \\ & (y^2, x_1^2, x_2^2, \dots, x_m^2), \\ & \dots, \\ & (y^n, x_1^n, x_2^n, \dots, x_m^n). \end{aligned} \quad (2.1)$$

Здесь и везде далее

n — количество наблюдений,

m — количество факторов.

Задача состоит в том, чтобы по данным наблюдениям подобрать аналитически описываемую функцию $\hat{Y} = f(X_1, X_2, \dots, X_m)$, которая в некотором смысле наилучшим образом описывает результаты наблюдений.

Для каждого i -го наблюдения можно записать: $y^i = \hat{y}^i + \varepsilon^i$,

где y^i — фактическое значение результата,

$\hat{y}^i$ — теоретическое значение результата,

ε^i — остаточная величина, показывающая отклонение фактического значения результата от теоретического.

Таким образом, в общем виде уравнение связи этих показателей имеет вид

$$Y = f(X_1, X_2, \dots, X_m) + \varepsilon = \hat{Y} + \varepsilon, \quad (2.2)$$

где $Y = (y^1, y^2, \dots, y^n)$,

$\hat{Y} = (\hat{y}^1, \hat{y}^2, \dots, \hat{y}^n)$,

$\varepsilon = (\varepsilon^1, \varepsilon^2, \dots, \varepsilon^n)$.

В частных случаях, если $m = 1$, то регрессия называется **парной** или **простой**; если $m > 1$, то регрессия называется **множественной**.

Если

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon, \quad (2.3)$$

то регрессия называется **линейной**, а иначе — **нелинейной**.

Точность уравнения (2.2) тем выше, чем меньше значения ε^i для всей совокупности наблюдений. На точность величины ε^i влияют следующие факторы:

- неправильный выбор функции f ;
- неточность измерения значений переменных x_i и y_i ;
- выборочный характер наблюдений (2.1).

2.2. Оценка качества уравнения регрессии

Для численной оценки вероятностных зависимостей случайных величин в таком случае используются

- индекс корреляции,
- индекс детерминации,
- скорректированный индекс корреляции,
- скорректированный индекс детерминации,
- средняя ошибка аппроксимации.

Обратимся к изучению этих оценок.

На основе этих наблюдений, как показано выше, общая дисперсия случайной величины Y определяется по формуле:

$$\sigma_{Y \text{ общ}}^2 = \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}.$$

Определим также дисперсию, объясненную дисперсией

$$\sigma_{Y \text{ регр}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2}{n},$$

и остаточную дисперсию

$$\sigma_{Y \text{ ост}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}.$$

Можно доказать, что

$$\sigma_{Y \text{ общ}}^2 = \sigma_{Y \text{ регр}}^2 + \sigma_{Y \text{ ост}}^2.$$

Индекс корреляции определяется по формуле:

$$r_{YX_1X_2...X_m} = r = \sqrt{1 - \frac{\sigma_{\bar{Y}}^2_{ост}}{\sigma_{\bar{Y}}^2_{общ}}} = \sqrt{\frac{\sigma_{\bar{Y}}^2_{регр}}{\sigma_{\bar{Y}}^2_{общ}}},$$

индекс детерминации –

$$r_{YX_1X_2...X_m}^2 = r^2 = 1 - \frac{\sigma_{\bar{Y}}^2_{ост}}{\sigma_{\bar{Y}}^2_{общ}} = \frac{\sigma_{\bar{Y}}^2_{регр}}{\sigma_{\bar{Y}}^2_{общ}}.$$

Уравнение регрессии называется значимым, если оно правильно отражает результаты наблюдений. Для проверки значимости уравнения регрессии используется F-критерий Фишера.

Также может быть оценена значимость отдельных факторов по t-критерию Стьюдента.

С этой целью по специальным статистическим таблицам рассчитывается уровень значимости α , который показывает вероятность того, что уравнение регрессии или фактор незначимы. Считается, что если $\alpha < 0,05$, то уравнение или фактор значимы, а иначе незначимы.

Замечания. Для линейных функциональных зависимостей индекс детерминации численно равен коэффициенту детерминации и характеризует долю случайной величины Y , объясненной уравнением регрессии к общей дисперсии этой величины; индекс корреляции равен модулю коэффициента корреляции.

В случае если число переменных велико или функциональная зависимость значительно отличается от линейной, вместо r и r^2 целесообразно использовать

скорректированный индекс корреляции

$$\tilde{r}_{YX_1X_2...X_m} = \tilde{r} = \sqrt{\max\left(0; 1 - (1 - r^2) \frac{n-1}{n-m-1}\right)},$$

скорректированный индекс детерминации

$$\tilde{r}_{YX_1X_2...X_m}^2 = \tilde{r}^2 = \max\left(0; 1 - (1 - r^2) \frac{n-1}{n-m-1}\right),$$

где n — число наблюдений,

m — число коэффициентов перед переменными в уравнении регрессии (для линейной регрессии это число равно числу переменных).

Правильность подбора функциональной зависимости (1.3) на основе выборки (1.4) определяется с помощью **средней ошибки аппроксимации**

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

Считается, что ошибка аппроксимации в пределах 5-7% свидетельствует о хорошем подборе функциональной зависимости (2.3).

2.3. Выбор вида уравнения парной регрессии

Рассмотрим способ определения вида функции f на примере наиболее важного частного вида парной регрессии — линейной парной регрессии $Y = a + bX + \varepsilon$.

Спецификация моделей парной регрессии сводится к выбору вида функции $f(X)$, т. к. изменение состава факторов привело бы к появлению модели множественной регрессии. Выбор вида функции может быть осуществлен тремя методами:

- ✓ теоретическим,
- ✓ графическим,
- ✓ экспериментальным.

Обратимся к рассмотрению этих методов.

Теоретический метод основан на качественном анализе взаимосвязи между переменными X и Y .

В таком случае функциональная зависимость между X и Y приобретает вид:

$$Y = a + bX + \varepsilon,$$

где a и b — числовые коэффициенты, которые необходимо оценить.

Графический метод основан на построении корреляционного поля или эмпирической линии регрессии.

Корреляционное поле — точечный график в системе координат XOY , в котором каждая точка соответствует одному наблюдению (рис. 2.1).

Эмпирическая линия регрессии графически представляет собой ломанную, соединяющую точки корреляционного поля (рис. 2.2).

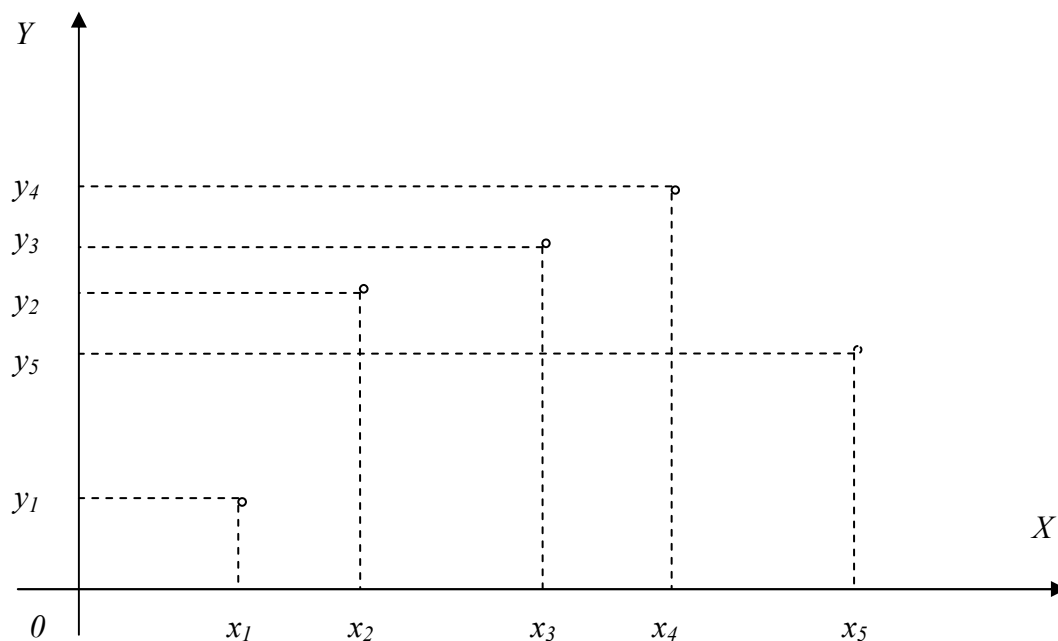


Рис. 2.1. Пример корреляционного поля, построенного по 5 наблюдениям

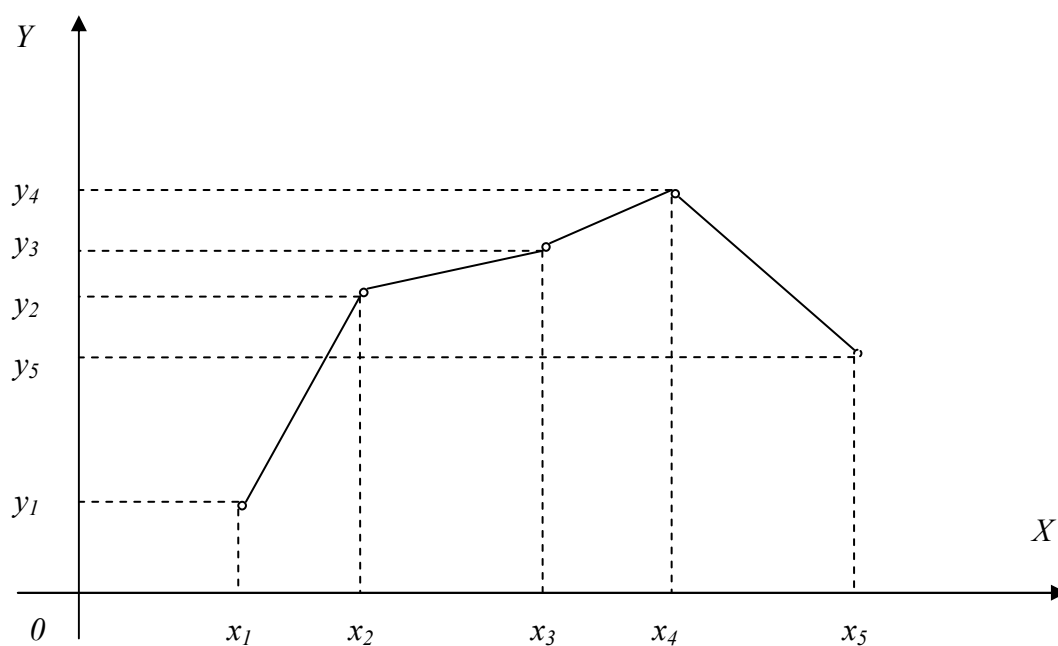


Рис. 2.2. Пример эмпирической линии регрессии, построенной по 5 наблюдениям

Графический метод предполагает, что исследователь, пользуясь полученными графическими представлениями, подберет такую функциональную зависимость, которая в наибольшей степени соответствует наблюдениям. Наиболее часто в парной регрессии используют следующие функциональные зависимости (рис. 2.3 — 2.8).

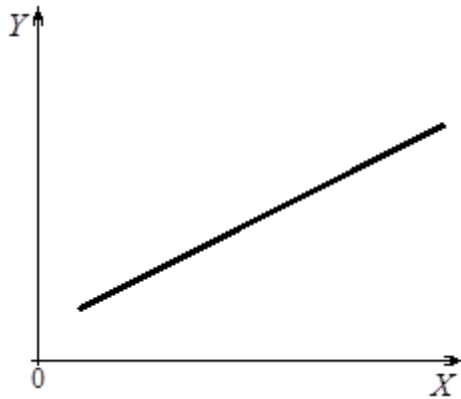


Рис. 2.3. Линейная — $\hat{Y} = a + b \cdot X$

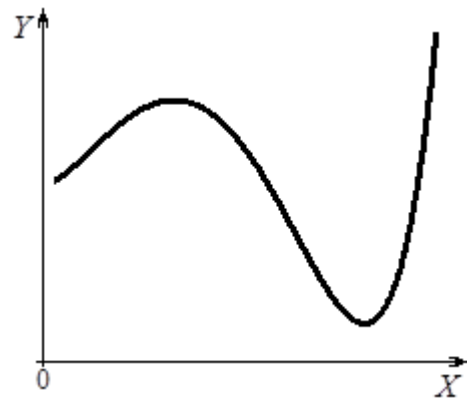


Рис. 2.4. Полиномиальная — $\hat{Y} = a + b_1 \cdot X + \dots + b_k \cdot X^k, k \geq 2$

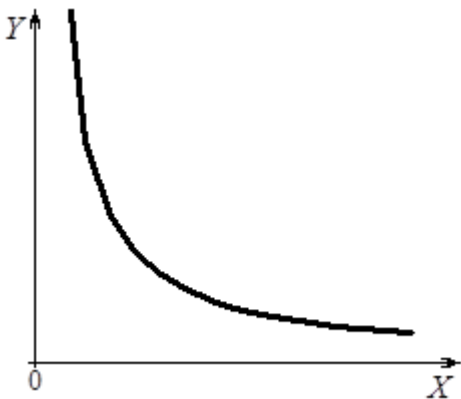


Рис. 2.5. Гиперболическая — $\hat{Y} = a + \frac{b}{X}$

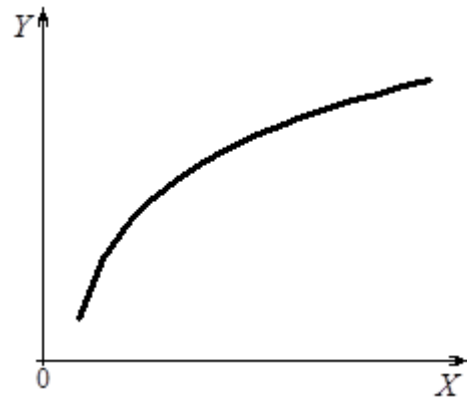


Рис. 2.6. Полулогарифмическая — $\hat{Y} = a + b \cdot \ln X$

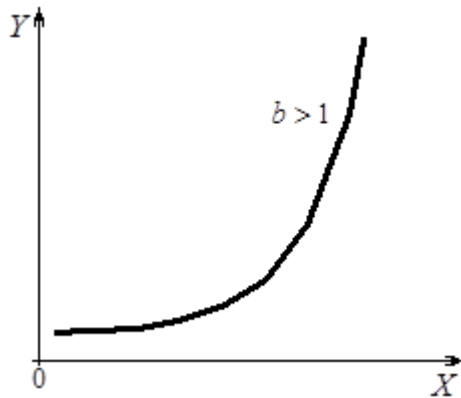


Рис. 2.7. Показательная — $\hat{Y} = a \cdot b^X$

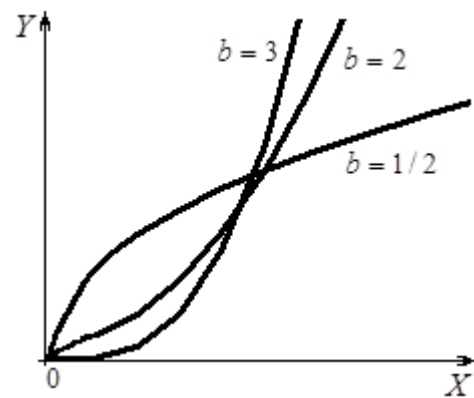


Рис. 2.8. Степенная — $\hat{Y} = a \cdot X^b$

Методы оценки числовых параметров a и b будут рассмотрены далее.

Экспериментальный метод заключается в следующем:

- строятся несколько уравнений регрессии (например, всех перечисленных выше типов),
- для каждого уравнения вычисляются показатели качества,
- выбирается функция $f(X)$, имеющая лучшие показатели качества.

Рассмотрим перечисленные задачи подробнее.

Для построения нелинейных уравнений регрессии обычно первоначально осуществляют их линеаризацию, т.е. приведение к линейному виду. Все уравнения делятся на два класса.

К первому классу относятся уравнения, линейные относительно параметров a и b . Их линеаризация может быть осуществлена заменой переменных (полиномиальная, гиперболическая и полулогарифмическая функции).

Пример. Пусть функциональная зависимость является гиперболической: $\hat{Y} = a + \frac{b}{X}$. Произведем замену $Z = \frac{1}{X}$. Тогда функциональная зависимость станет линейной: $\hat{Y} = a + bZ$.

Ко второму классу относятся уравнения, нелинейные относительно параметров a и b . Их линеаризация может быть осуществлена нелинейным функциональным преобразованием (показательная и степенная функции).

Пример. Пусть функциональная зависимость является показательной: $\hat{Y} = ab^X$. Произведем логарифмирование: $\ln \hat{Y} = \ln a + (\ln b)X$. Произведем замену $Z = \ln \hat{Y}$. Тогда функциональная зависимость станет линейной: $Z = \ln a + (\ln b)X$.

В качестве показателей качества выбирают показатели, перечисленные в разделе 1. При этом для уравнений линейных и нелинейных первого класса используются коэффициенты или индексы корреляции и детерминации, а для нелинейных второго класса — поправленные индексы корреляции и детерминации.

2.4. Оценка параметров уравнения линейной парной регрессии

Рассмотрим пример: зависимость между численностью сотрудников x_i в различных отделах и количеством преступлений y_i в подотчетных данным отделам районах.

№ отдела	1	2	3	4	5	6	7	8	9	10
x_i	6	7	7	7	5	8	4	6	9	10
y_i	62	77	67	69	72	79	41	58	93	97

Изобразим указанные значения с помощью графика (рис. 2.9).

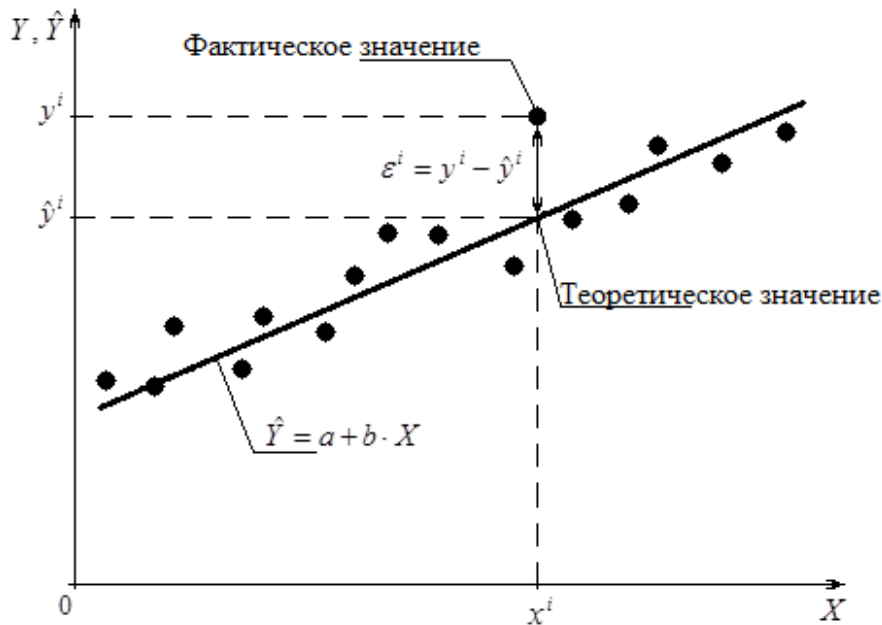


Рис. 2.9. Графическое изображение наблюдений

Легко видеть, что указанные точки наиболее близко расположены от прямой, указанной на рисунке.

Поэтому зависимость целесообразно искать в виде

$$Y = a + bX + \varepsilon. \quad (2.4)$$

Уравнение вида (2.4) называется уравнением линейной парной регрессии. Определение функции регрессии в этом случае сводится к нахождению численных значений параметров a и b . Для нахождения этих параметров часто используют метод наименьших квадратов (МНК).

В соответствии с МНК параметры a и b выбирают так, чтобы сумма квадратов отклонений фактических значений результата от теоретических ε_i было минимальным.

Геометрически эта задача представляет построение прямой с уравнением $Y = a + bX$ так, чтобы она максимально близко проходила от точек, задаваемых наблюдениями (рис. 2.8).

В соответствии с изученными выше свойствами случайных величин:

$$b = \frac{\text{cov}(X, Y)}{\sigma_X^2},$$

где $\text{cov}(X, Y)$ — ковариация случайных величин X и Y ,

σ_X^2 — дисперсия случайной величины X .

Параметр b называется **коэффициентом регрессии**. Он имеет следующий содержательный смысл:

коэффициент регрессии показывает, на сколько единиц в среднем изменится результат Y , если фактор X изменится на одну единицу.

Действительно, преобразуем уравнение (2.4) следующим образом:

$$Y_1 = a + b(X + 1) + \varepsilon = a + bX + \varepsilon + b.$$

Таким образом, $Y_1 = Y + b$.

Параметр a , как правило, содержательного смысла не имеет.

3. МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

3.1. Структура временного ряда

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени.

Значение ряда, относящееся к моменту времени t , обозначается y_t и называется уровнем ряда. Последовательность всех уровней ряда будем обозначать $Y = (y_1, y_2, \dots, y_n)$.

Изучим факторы, влияющие на формирование временного ряда.

С одной стороны, большинство показателей временных рядов имеют возрастающую или убывающую тенденцию (тренд), характеризующую совокупное долговременное воздействие множества факторов.

С другой стороны, показатели часто носят сезонный характер, т. к. деятельность человека зависит от времени года.

Однако, в большинстве рядов наряду с указанными факторами действуют еще и иные, носящие случайный или непостоянный характер факторы.

Таким образом, все факторы, на основе которых формируется временной ряд, подразделяются на три группы:

- факторы, формирующие тенденцию ряда,
- факторы, формирующие циклические (сезонные) колебания ряда,
- факторы, носящие случайный или непостоянный характер.

Каждая из описанных компонент может отсутствовать.

Последовательность значений уровней ряда, полученных в результате воздействия факторов, формирующих тенденцию, будем обозначать — $T = (t_1, t_2, \dots, t_n)$;

последовательность значений уровней ряда, полученных в результате воздействия факторов, формирующих сезонные колебания — $S = (s_1, s_2, \dots, s_n)$;

последовательность значений уровней ряда, полученных в результате воздействия факторов, которые носят случайный характер — $E = (e_1, e_2, \dots, e_n)$.

Таким образом, временной ряд является функцией трех компонент:

- 1) тренда — T ,
- 2) сезонной — S ,
- 3) случайной — E

$$Y = f(T, S, E). \quad (3.1)$$

3.2. Выявление компонент временного ряда

Обычно моделирование временных рядов осуществляется в следующей последовательности:

- 1) выявление наличия сезонной компоненты и в случае ее наличия — числа уровней, относящихся к одному периоду;
- 2) определения типа модели временного ряда — аддитивной или мультипликативной;
- 3) нахождение количественного выражения сезонной компоненты;
- 4) нахождение количественного выражения трендовой компоненты;
- 5) нахождение количественного выражения случайной компоненты;

6) оценка качества построенной модели временного ряда.

Основная задача исследования временного ряда состоит в определении типа его модели и количественном выражении каждой из компонент в интересах прогнозирования будущих значений ряда или исследования взаимосвязи двух и более временных рядов.

Первой задачей, которую необходимо решить при выявлении структуры временного ряда (3.1), является определение вида функциональной зависимости $f(T, S, E)$. Если функциональная зависимость (3.1) представляет собой сумму перечисленных компонент — $Y = T + S + E$, то она называется **аддитивной** моделью временного ряда, если произведение — $Y = T \cdot S \cdot E$, то — **мультипликативной** моделью временного ряда. Определение типа модели основано на использовании функции автокорреляции.

Обозначим $r_k = \frac{\text{cov}(Y, Y_k)}{\sigma_Y \sigma_{Y_k}}$ — коэффициент корреляции исходного ряда (3.1) и

ряда Y_k , полученного из исходного сдвигом на k уровней. Непосредственной проверкой определим значение $r^* = \max |r_k|$. Если r^* близок к 1, то это свидетельствует о наличии автокорреляции уровней ряда [3, 4].

Обозначим $k^* = \text{Arg max } |r_k|$. Если $k^* > 1$, то можно сделать вывод о наличии циклической компоненты с периодом, включающим k^* уровней ряда; если $k^* = 1$, то ряд включает только трендовую компоненту. Разобьем множество уровней ряда на периоды изменения циклической компоненты, т. е. на непересекающиеся отрезки, включающие k^* уровней ряда, и найдем амплитуду изменения уровней ряда в каждом. Заметим, что если $k^* = 1$, то последовательность амплитуд является монотонной (безразлично — возрастающей или убывающей), тогда целесообразно осуществлять построение мультипликативной модели, иначе — аддитивной [4].

Для нахождения сезонной компоненты может быть использован метод скользящей средней. Его суть заключается в следующем.

Первоначально находятся скользящие средние по формуле

$$y_{\text{скол}}^i = \begin{cases} \frac{1}{2} y^{i-\frac{k^*}{2}} + y^{i-\frac{k^*}{2}+1} + \dots + y^{i+\frac{k^*}{2}-1} + \frac{1}{2} y^{i+\frac{k^*}{2}}, & \text{если } k^* \text{ четно,} \\ y^{i-\frac{k^*-1}{2}} + y^{i-\frac{k^*}{2}+1} + \dots + y^{i+\frac{k^*}{2}-1} + y^{i+\frac{k^*-1}{2}}, & \text{если } k^* \text{ нечетно.} \end{cases}$$

После этого находятся средние значения скользящих средних для каждого уровня одного периода:

$$\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{k^*}.$$

Полученные значения представляют собой оценки циклической компоненты. Для нахождения точных значений необходимо преобразовать эти оценки в оценки

$$\tilde{y}'_1, \tilde{y}'_2, \dots, \tilde{y}'_{k^*} \quad (3.2)$$

так, чтобы выполнялось свойство $\sum_{i=1}^{k^*} \tilde{y}'_i = 0$ - для аддитивной модели и $\prod_{i=1}^{k^*} \tilde{y}'_i = 1$ — для мультипликативной модели. Значения (3.2) могут быть найдены по формулам

$$\tilde{y}'_i = \tilde{y}_i - \frac{\sum_{i=1}^{k^*} \tilde{y}_i}{k^*} \text{ — для аддитивной модели и } \tilde{y}'_i = \frac{\tilde{y}_i}{\sqrt[k^*]{\prod_{i=1}^{k^*} \tilde{y}_i}} \text{ — для мультипликативной}$$

модели. После этого значения циклической компоненты C_t находятся по формуле $c'_t = \tilde{y}'_{t(\text{mod } k^*)}$, где $t(\text{mod } k^*)$ — остаток от деления t на k^* .

После нахождения циклических компонент находится трендовая компонента. Для этого из уровней ряда исключаются значения циклических компонент по формулам

$$\hat{y}^t = y^t - \sum_{i=1}^m c_i^t \text{ — для аддитивной модели,}$$

$$\hat{y}^t = \frac{y^t}{\prod_{i=1}^m c_i^t} \text{ — для мультипликативной модели.}$$

Нахождение трендовой компоненты осуществляется методом аналитического выравнивания, под которым понимается построение уравнения парной регрессии значений $\hat{Y} = (\hat{y}^1, \hat{y}^2, \dots, \hat{y}^n)$ по номерам уровней ряда. Как правило, при анализе структуры временных рядов статистики достаточно использовать линейную регрессию. При этом могут быть выведены значения трендовой компоненты $T = (t^1, t^2, \dots, t^n)$ и найдены значения случайной компоненты по формулам

$$e^t = \hat{y}^t - t^t \text{ — для аддитивной модели;}$$

$$e^t = \frac{\hat{y}^t}{t^t} \text{ — для мультипликативной модели.}$$

4. ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Пример 1. При изучении дисциплины «НХКБ» студент Думачёв получил следующие оценки: 3, 2, 5, 4, 2, 4, 3, 5, 5, 2, 3, 3, 4, 4, 5, 4, 2, 4, 2, 4, 5, 5, 2, 3, 4, 2, 5, 4, 5, 4, 2, 4, 3, 3, 5, 3, 2, 4, 5, 2, 5.

Необходимо:

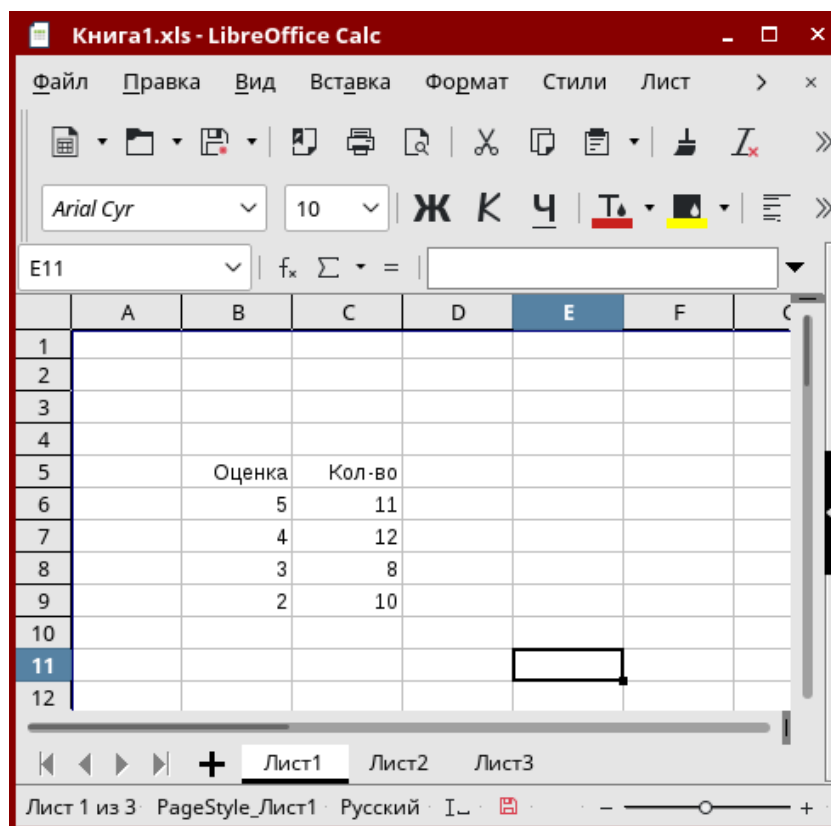
1. Построить гистограмму (график) распределения оценок.
2. Определить средний бал студента Думачёва, найти разброс оценок около среднего бала.
3. Сделать вывод о качестве освоения Думачёвым дисциплины «НХКБ».

Решение.

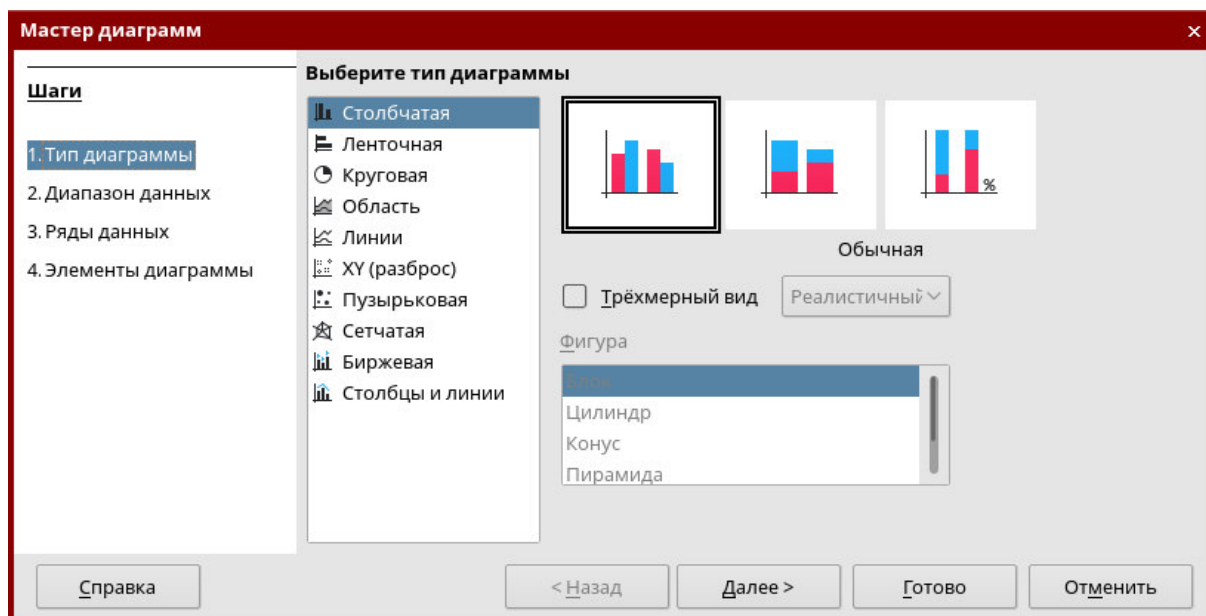
1. Из полученных студентом Думачёвым оценок составим вариационный ряд 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, значения которого сведём в таблицу.

Оценка (качественный показатель)	5	4	3	2
Количество (количественный показатель)	11	12	8	10

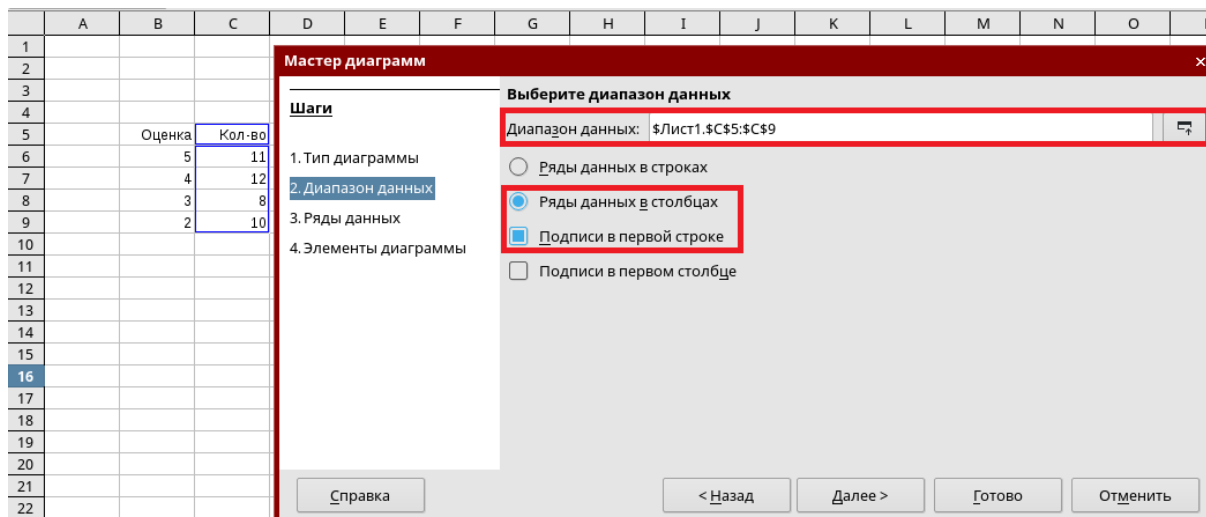
Для лучшего восприятия и наглядности полученной информации построим гистограмму распределения полученных оценок. Запускаем программу Пуск → Офис LibreOffice → Создать: Таблицу Calc. Вбиваем значения из таблицы и вызываем «Мастер диаграмм» с помощью пиктограммы или «Вставка — Диаграмма».



Выбираем «Тип диаграммы → Столбчатая» и нажимаем «Далее».



В диапазон данных вводим номера ячеек, в которых стоят значения количества оценок и заголовков, ставим галочки в полях «Ряды данных в столбцах» и «Подписи в первой строке». Нажимаем «Далее».



На шаге «Ряды данных» в поле «Категории» вводим номера ячеек с оценками. Нажимаем «Далее».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3																
4																
5		Оценка	Кол-во													
6		5	11													
7		4	12													
8		3	8													
9		2	10													
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																

Мастер диаграмм

Шаги

1. Тип диаграммы
2. Диапазон данных
- 3. Ряды данных**
4. Элементы диаграммы

Ряды данных: Кол-во

Диапазоны данных:

Диапазон для %VALUETYPE

Категории

\$Лист1.\$B\$6:\$B\$9

Добавить

Удалить

Настроить диапазоны данных для каждого ряда данных

Справка

< Назад

Далее >

Готово

Отменить

На последнем шаге мастера диаграмм заполняем поля «Заглавие» и «Ось X».

Мастер диаграмм

Шаги

1. Тип диаграммы
2. Диапазон данных
3. Ряды данных
- 4. Элементы диаграммы**

Укажите заглавия, легенду и параметры сетки

Заглавие Успеваемость студента

Подзаголовок

Ось X Оценки

Ось Y

Ось Z

Отображать сетки

☐ Ось X ☒ Ось Y ☐ Ось Z

☒ Показать легенду

☐ Слева

☒ Справа

☐ Сверху

☐ Снизу

Справка

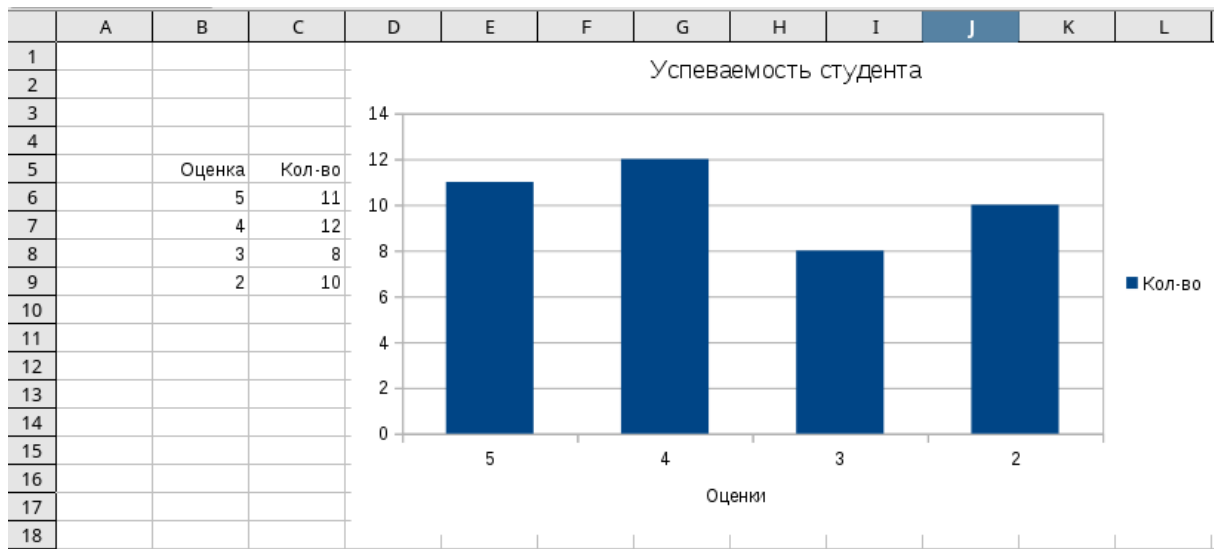
< Назад

Далее >

Готово

Отменить

Нажимаем «Готово», получаем желаемый результат.



2. Оценки, полученные Думачёвым, являются случайными величинами, так как могут принимать различные значения в зависимости от того, подготовился студент к занятию или нет или по каким-то другим причинам. Поэтому, для изучения успеваемости воспользуемся элементами математической статистики.

Для вычисления среднего бала студента, а также разброса значений около среднего можно воспользоваться формулами

$$x_{\text{ср}} = \frac{\sum_{i=1}^m x_i \cdot n_i}{\sum_{i=1}^m n_i}, \quad D(X) = M(X^2) - M^2(X), \quad \sigma = \sqrt{D(X)},$$

где $x_{\text{ср}}$ — средний бал студента (среднее значение);

x_i — качественный показатель (т.е. сама оценка);

n_i — количественный показатель;

m — общее количество оценок;

$D(X)$ — дисперсия;

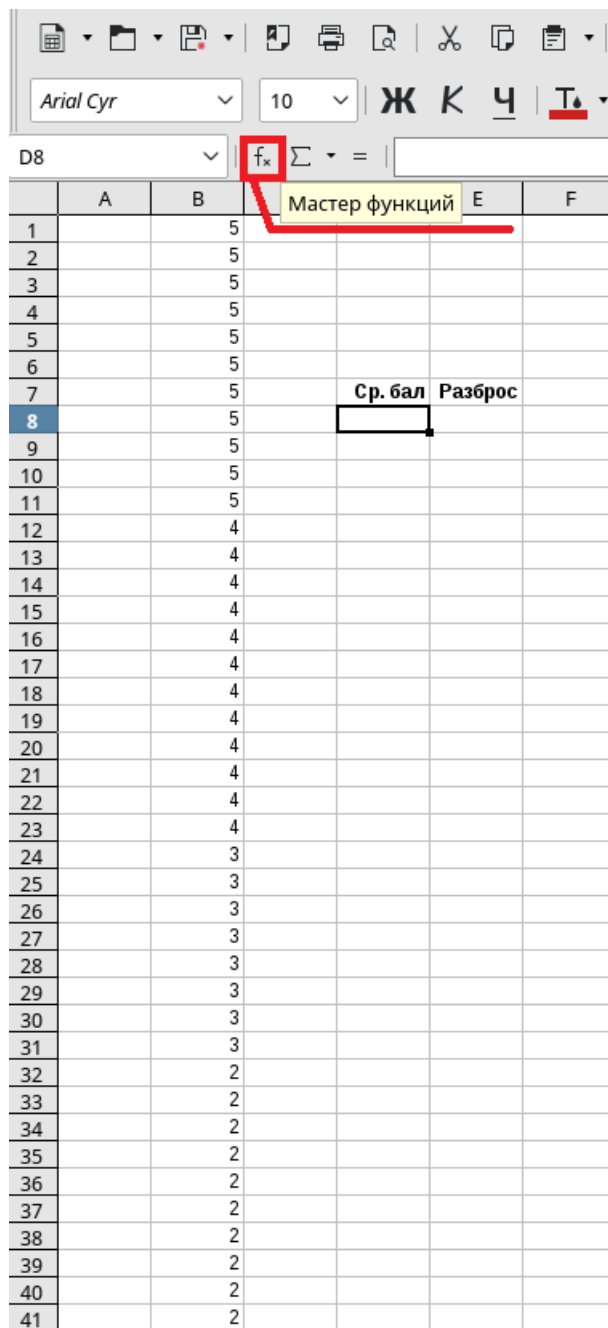
$M(X)$ — средний бал,

$M(X^2)$ — второй начальный момент или средний бал, вычисленный при возведении x_i в квадрат;

σ — среднеквадратическое отклонение случайной величины около среднего значения, т. е. то, что и нужно найти.

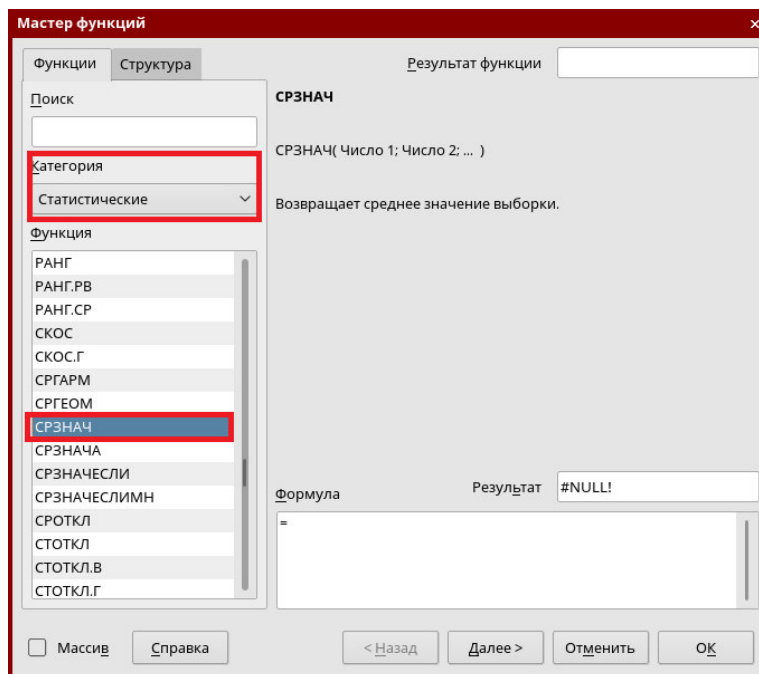
Однако, данные формулы и те, которые будут применяться в дальнейшем для расчета различных характеристик, тяжелы для восприятия и запоминания. Поэтому воспользуемся более простым способом для нахождения нужных значений.

Вводим в Calc оценки, полученные студентом (в любом виде: в столбик, в строку, в произвольной последовательности, в виде вариационного ряда т. п.). Для удобства подпишем ячейки как средний балл и разброс значений.

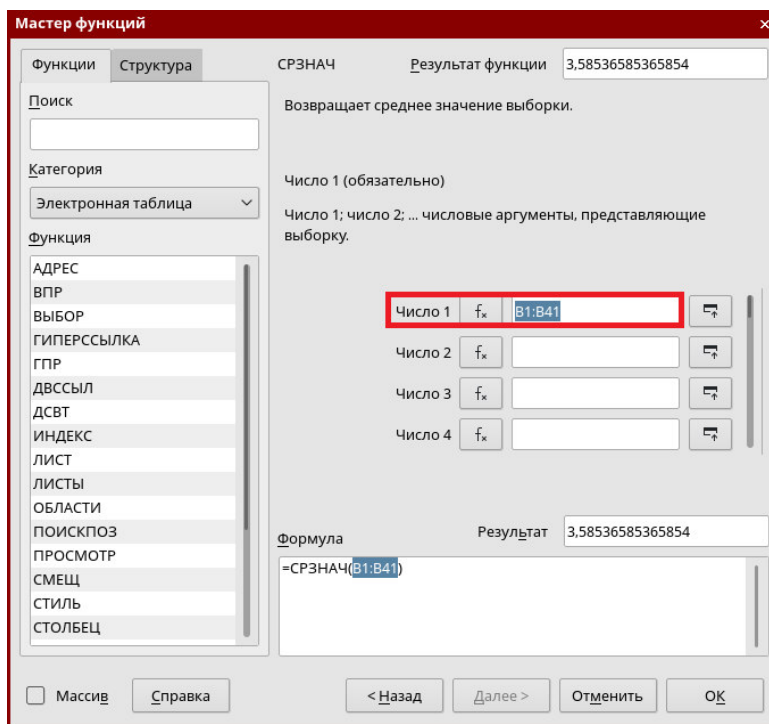


	A	B	E	F
1		5		
2		5		
3		5		
4		5		
5		5		
6		5		
7		5		
8		5	Ср. бал	Разброс
9		5		
10		5		
11		5		
12		4		
13		4		
14		4		
15		4		
16		4		
17		4		
18		4		
19		4		
20		4		
21		4		
22		4		
23		4		
24		3		
25		3		
26		3		
27		3		
28		3		
29		3		
30		3		
31		3		
32		2		
33		2		
34		2		
35		2		
36		2		
37		2		
38		2		
39		2		
40		2		
41		2		

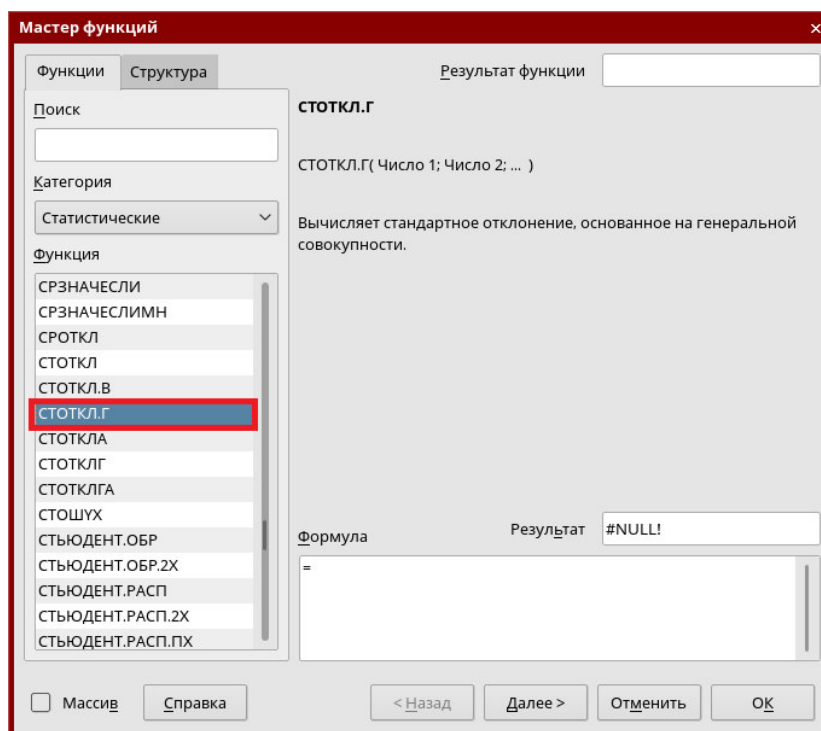
Для расчета среднего значения ставим курсор под «Ср. бал» (т. е. в нужном нам месте), подгружаем «Мастер функций»: «Вставка — Функция» или воспользуемся пиктограммой «Мастер функций». Далее в поле «Категория» выбираем «Статистические». В поле «Функция» находим «СРЗНАЧ».



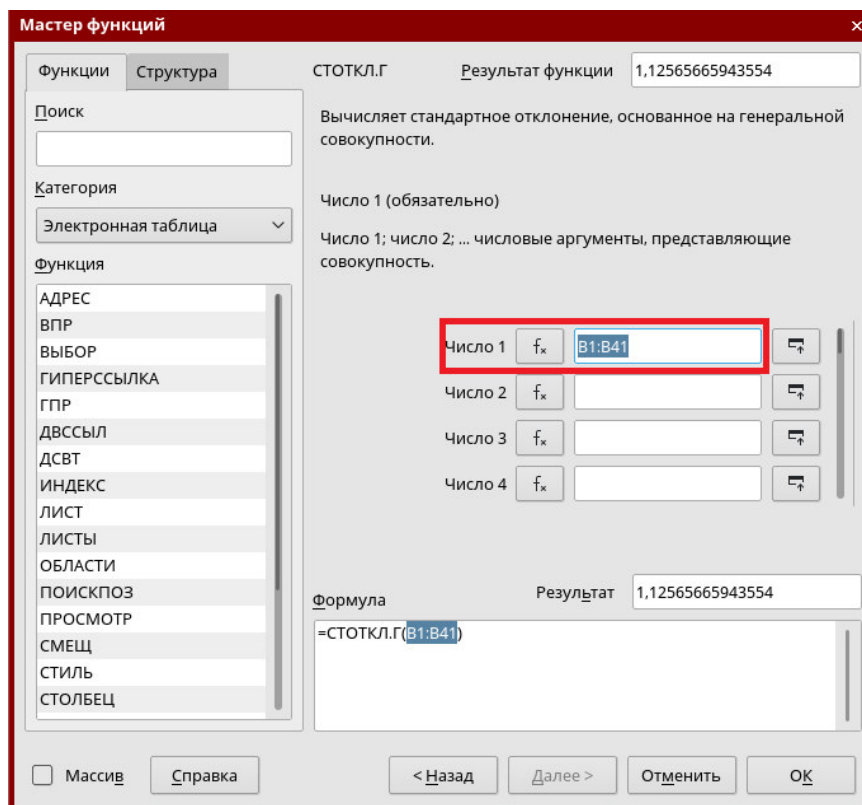
Нажимаем «Далее». В поле «Число 1» вводим диапазон ячеек с нашими данными. Нажимаем «ОК».



Для расчета разброса значений около среднего значения ставим курсор под «Разброс» (т. е. в нужном нам месте) подгружаем «Мастер функций»: «Вставка — Функция». Далее в поле «Категория» выбираем «Статистические». В поле «Функция» находим «СТОТКЛ.Г».



Нажимаем «Далее». В поле «Число 1» вводим диапазон ячеек с нашими данными. Нажимаем «ОК».



В результате получаем нужные расчетные значения.

	A	B	C	D	E	F	G	H	I	J
1		5								
2		5								
3		5								
4		5								
5		5								
6		5								
7		5								
8		5								
9		5								
10		5								
11		5								
12		4								
13		4								
14		4								
15		4								
16		4								
17		4								
18		4								
19		4								
20		4								
21		4								
22		4								
23		4								
24		3								
25		3								
26		3								
27		3								
28		3								
29		3								
30		3								
31		3								
32		2								
33		2								
34		2								
35		2								
36		2								
37		2								
38		2								
39		2								
40		2								
41		2								
42										

Ср. бал Разброс
3,585366 1,125657

3. Таким образом, средний бал студента Думачёва при изучении дисциплины «НХКБ» составил 3,58, при этом разброс значений (оценок) около среднего составил 1,12.

Данный средний бал показывает усредненную оценку качества полученных знаний при изучении дисциплины. Разброс значений около среднего, показывает диапазон изменения случайной величины, в данном случае оценок. То есть, студент учился на 3,5, однако его оценки находятся в интервале от $3,58 - 1,12 = 2,36$ до $3,58 + 1,12 = 4,7$. Данный факт показывает, что студент — способный ученик, но он либо халатно относился к учебе, что скорее всего, либо ему при ответе попадались сложные вопросы.

Пусть, другой студент, имеет средний бал 3,5 и разброс значений около данной величины 0,1. Очевидно, что он учится стабильно, не допускает срывов в учебе.

Какого студента назвать «лучшим»? Что лучше: не очень высокий результат, но стабильность, или хорошие показатели, но ненадежность?

Как видно из примера, любое событие можно описать количественными характеристиками, по которым можно охарактеризовать произошедшее событие. Однако мало знать только среднее значение, необходимо учитывать и разброс значений около среднего.

Пример 2. В городе N-ске были получены статистические данные по 19 опорным пунктам участковых инспекторов полиции за отчетный период, которые сведены в таблицу.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	54	25500	28	51
2.	65	27900	35	60
3.	85	32000	45	81
4.	68	27800	36	62
5.	95	33500	50	92
6.	53	24600	29	46
7.	67	31200	36	62
8.	61	32400	33	59
9.	59	48500	32	53
10.	74	37900	39	73
11.	58	24800	31	50
12.	43	23700	24	40
13.	76	46700	40	71
14.	80	41450	42	76
15.	72	34200	38	70
16.	93	31500	49	85
17.	56	46900	30	55
18.	74	33200	39	71
19.	34	12000	19	29

Необходимо:

1. Определить существование зависимости количества преступлений от численности населения, количества осужденных и количества раскрытых преступлений по городу в целом.

2. Сделать прогноз оценки количества преступлений на 5 участке, если количество осужденных уменьшится на 10, а количество раскрытых преступлений увеличится на 15.

Решение.

Для решения данной задачи воспользуемся редактором Calc. Заносим данные в виде таблицы.

	A	B	C	D	E	F	G
1							
2		№ п/п	Кол-во преступлений	Численность населения	Кол-во осужденных	Кол-во раскрытых преступлений	
3		1.	54	25500	28	51	
4		2.	65	27900	35	60	
5		3.	85	32000	45	81	
6		4.	68	27800	36	62	
7		5.	95	33500	50	92	
8		6.	53	24600	29	46	
9		7.	67	31200	36	62	
10		8.	61	32400	33	59	
11		9.	59	48500	32	53	
12		10.	74	37900	39	73	
13		11.	58	24800	31	50	
14		12.	43	23700	24	40	
15		13.	76	46700	40	71	
16		14.	80	41450	42	76	
17		15.	72	34200	38	70	
18		16.	93	31500	49	85	
19		17.	56	46900	30	55	
20		18.	74	33200	39	71	
21		19.	34	12000	19	29	
22							

1. В меню «Данные» выбираем «Статистика» → «Регрессия...».

В поле «Диапазон независимых переменных (X)» вводим ячейки с влияющими факторами на изучаемый параметр — численность населения, количество осужденных, количество раскрытых преступлений.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2			№ п/п	Кол-во преступлений	Численность населения	Кол-во осужденных	Кол-во раскрытых преступлений								
3			1.	54	25500	28	51								
4			2.	65	27900	35	60								
5			3.	85	32000	45	81								
6			4.	68	27800	36	62								
7			5.	95	33500	50	92								
8			6.	53	24600	29	46								
9			7.	67	31200	36	62								
10			8.	61	32400	33	59								
11			9.	59	48500	32	53								
12			10.	74	37900	39	73								
13			11.	58	24800	31	50								
14			12.	43	23700	24	40								
15			13.	76	46700	40	71								
16			14.	80	41450	42	76								
17			15.	72	34200	38	70								
18			16.	93	31500	49	85								
19			17.	56	46900	30	55								
20			18.	74	33200	39	71								
21			19.	34	12000	19	29								
22															
23															
24															
25															
26															
27															

Регрессия

Данные

Диапазон независимых переменных (X):

Диапазон зависимой переменной (Y):

☐ Оба диапазона X и Y с подписями

Результат в:

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия:

☐ Вычислить остатки ☐ Свободный член равен 0

Неверный диапазон зависимых переменных.

В поле «Диапазон зависимой переменной (Y)» вводим номера ячеек, содержащие изучаемый параметр — количество преступлений.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2			№ п/п	Кол-во преступлений	Численность населения	Кол-во осужденных	Кол-во раскрытых преступлений								
3			1.	54	25500	28	51								
4			2.	65	27900	35	60								
5			3.	85	32000	45	81								
6			4.	68	27800	36	62								
7			5.	95	33500	50	92								
8			6.	53	24600	29	46								
9			7.	67	31200	36	62								
10			8.	61	32400	33	59								
11			9.	59	48500	32	53								
12			10.	74	37900	39	73								
13			11.	58	24800	31	50								
14			12.	43	23700	24	40								
15			13.	76	46700	40	71								
16			14.	80	41450	42	76								
17			15.	72	34200	38	70								
18			16.	93	31500	49	85								
19			17.	56	46900	30	55								
20			18.	74	33200	39	71								
21			19.	34	12000	19	29								
22															
23															
24															
25															
26															

Регрессия

Данные

Диапазон независимых переменных (X):

Диапазон зависимой переменной (Y):

☐ Оба диапазона X и Y с подписями

Результат в:

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия:

☐ Вычислить остатки ☐ Свободный член равен 0

Недопустимый выходной адрес.

Рассмотрим полученные результаты.

	G	H	I	J	K	L	M	N	O
3		Регрессия							
4		Регрессионная модель	Линейная						
5									
6		Необработанный вывод ЛИНЕЙН							
7		0,167918250239648	1,649043430046	-9,02417899856224E-06	-2,089406413				
8		0,069576342881704	0,1358391952218	2,06763399153625E-05	0,99716092971				
9		0,998447519192464	0,6818556644181	#Н/Д	#Н/Д				
8		3215,6517309124	15	#Н/Д	#Н/Д				
9		4485,13135595141	6,9739072064853	#Н/Д	#Н/Д				
10									
11		Статистика регрессии							
14		R^2	0,9984475191925						
15		Среднеквадратическое отклонение	0,6818556644181						
15		Число переменных X	3						
14		Наблюдения	19						
15		Скорректированный R^2	0,998137023031						
		Дисперсионный анализ (ANOVA)							
21			df	SS	MS	F	Значимость F		
22		Регрессия	3	4485,13135595141	1495,04378532	3215,65173	2,7744739E-21		
17		Остатки	15	6,97390720648533	0,4649271471				
		Итог	18	4492,10526315789					
18									
		Уровень доверия	0,95						
19									
27									
28			Коэффициенты	Среднеквадратическое отклонение	t-статистика	P-значение	Нижний 95%	Верхний 95%	
29		Пересечение	-2,089406412995	0,997160929709104	-2,0953552739	0,05352078	-4,21480462347	0,0359917975	
30		X1	-9,024179E-06	2,06763399153625E-05	-0,4364495378	0,66872742	-5,3094754E-05	3,50464E-05	
31		X2	1,649043430046	0,135839195221813	12,1396731433	3,6898E-09	1,35950903912	1,938577821	
32		X3	0,1679182502396	0,069576342881704	2,41343886851	0,0290512	0,01961978582	0,3162167147	
33									

Значение множественного коэффициента регрессии должно быть близким к 1 («R^2» — выделено красным цветом). В противном случае, все имеющиеся входные статистические данные сфальсифицированы и не имеют ничего общего с реальной обстановкой.

Значение «Значимость F» (выделено зеленым цветом) дисперсионного анализа должно быть меньше 0,05, что было определено выше, и указывает на величину ошибки.

Оба этих параметра в нашем примере соответствуют указанным диапазонам, поэтому полученным результатам можно доверять с вероятностью 95%.

Значение «Пересечение» третьей таблицы соответствует количеству преступлений, переменная «X1» — численности населения, переменная «X2» — количество осужденных, переменная «X3» — количество раскрытых преступлений.

У данных переменных величина «P-Значение» (выделено желтым цветом) должна быть меньше 0,05. Если величина значение «P-Значение» превышает 0,05, то данная переменная не влияет на изучаемый параметр. В нашем случае «X2» и «X3» имеют «P-Значение» меньше 0,05, а «X1» — больше. Таким образом, количество осужденных и количество раскрытых преступлений влияют на общее количество преступлений, а численность населения — нет.

Удаляем столбец с численностью населения. И проводим с оставшимися данными те же действия. Если все параметры в допустимых пределах, то переходим к проверке данных и прогнозированию (см. ниже).

	A	B	C	D	E	F
1						
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений	
3		1.	54	28	51	
4		2.	65	35	60	
5		3.	85	45	81	
6		4.	68	36	62	
7		5.	95	50	92	
8		6.	53	29	46	
9		7.	67	36	62	
10		8.	61	33	59	
11		9.	59	32	53	
12		10.	74	39	73	
13		11.	58	31	50	
14		12.	43	24	40	
15		13.	76	40	71	
16		14.	80	42	76	
17		15.	72	38	70	
18		16.	93	49	85	
19		17.	56	30	55	
20		18.	74	39	71	
21		19.	34	19	29	
22						

В меню «Данные» выбираем «Статистика» → «Регрессия...». В поле «Диапазон независимых переменных (X)» вводим ячейки с влияющими факторами на изучаемый параметр — количество осужденных, количество раскрытых преступлений.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений							
3		1.	54	28	51							
4		2.	65	35	60							
5		3.	85	45	81							
6		4.	68	36	62							
7		5.	95	50	92							
8		6.	53	29	46							
9		7.	67	36	62							
10		8.	61	33	59							
11		9.	59	32	53							
12		10.	74	39	73							
13		11.	58	31	50							
14		12.	43	24	40							
15		13.	76	40	71							
16		14.	80	42	76							
17		15.	72	38	70							
18		16.	93	49	85							
19		17.	56	30	55							
20		18.	74	39	71							
21		19.	34	19	29							
22												
23												
24												
25												
26												
27												

Регрессия

Данные

Диапазон независимых переменных (X): \$Лист1.\$D\$3:\$E\$21

Диапазон зависимой переменной (Y):

☐ Оба диапазона X и Y с подписями

Результат в:

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия: 0,95

☒ Вычислить остатки ☐ Свободный член равен 0

Неверный диапазон зависимых переменных.

Справка Отменить ОК

В поле «Диапазон зависимой переменной (Y)» вводим номера ячеек, содержащие изучаемый параметр — количество преступлений.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений							
3		1.	54	28	51							
4		2.	65	35	60							
5		3.	85	45	81							
6		4.	68	36	62							
7		5.	95	50	92							
8		6.	53	29	46							
9		7.	67	36	62							
10		8.	61	33	59							
11		9.	59	32	53							
12		10.	74	39	73							
13		11.	58	31	50							
14		12.	43	24	40							
15		13.	76	40	71							
16		14.	80	42	76							
17		15.	72	38	70							
18		16.	93	49	85							
19		17.	56	30	55							
20		18.	74	39	71							
21		19.	34	19	29							
22												
23												
24												
25												
26												
27												
28												

Регрессия

Данные

Диапазон независимых переменных (X): \$Лист1.\$D\$3:\$E\$21

Диапазон зависимой переменной (Y): \$Лист1.\$C\$3:\$C\$21

☐ Оба диапазона X и Y с подписями

Результат в:

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия: 0.95

☒ Вычислить остатки ☐ Свободный член равен 0

Недопустимый выходной адрес.

Справка Отменить ОК

Выведем расчетные значения на этот же лист, выбрав «Результаты в:» и указав ячейку (это левый верхний угол таблицы с результатами).

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений							
3		1.	54	28	51							
4		2.	65	35	60							
5		3.	85	45	81							
6		4.	68	36	62							
7		5.	95	50	92							
8		6.	53	29	46							
9		7.	67	36	62							
10		8.	61	33	59							
11		9.	59	32	53							
12		10.	74	39	73							
13		11.	58	31	50							
14		12.	43	24	40							
15		13.	76	40	71							
16		14.	80	42	76							
17		15.	72	38	70							
18		16.	93	49	85							
19		17.	56	30	55							
20		18.	74	39	71							
21		19.	34	19	29							
22												
23												
24												
25												
26												
27												
28												

Регрессия

Данные

Диапазон независимых переменных (X): \$Лист1.\$D\$3:\$E\$21

Диапазон зависимой переменной (Y): \$Лист1.\$C\$3:\$C\$21

☐ Оба диапазона X и Y с подписями

Результат в: \$G\$3

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия: 0.95

☐ Вычислить остатки ☐ Свободный член равен 0

Справка Отменить ОК

Нажимаем «ОК».

Таким образом, на изучаемый параметр (общее количество преступлений) влияют такие факторы как количество осужденных и количество раскрытых преступлений. Данные результаты получены с достоверностью 95%.

Теперь есть возможность делать прогноз по каждому из участков. Каково будет общее количество преступлений при изменении количества осужденных и количества раскрытых преступлений? Для этого рассмотрим предыдущие результаты.

Сначала сделаем проверку сделанных вычислений, чтобы убедиться в правильности расчета встроенного в Calc пакета анализа данных.

В третьей табличке во втором столбце находятся значения коэффициентов (выделены синим цветом). Это коэффициенты уравнения регрессии, которое в общем случае имеет вид:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + \dots + a_n \cdot x_n,$$

где y — изучаемый параметр;

a_0 — число, стоящее на пересечении «Коэффициенты» и «Y-пересечение»;

a_i — числа, стоящие на пересечении «Коэффициенты» и «Переменная X»;

$x_1, \dots, x_n$ — факторы, влияющие на изучаемый параметр.

	F	G	H	I	J	K	L	M	N
1									
2									
3		Регрессия							
4		Регрессионная модель	Линейная						
5		Необработанный вывод ЛИНЕЙН							
6									
7		0,158855339068342	1,66229790873236	-2,28702739628397					
8		0,06470373521014	0,129008125551406	0,865664125994555					
9		0,998427803907596	0,664382718271549	#Н/Д					
10		5080,42379054967	16	#Н/Д					
11		4485,04279281649	7,06247034140628	#Н/Д					
12									
13		Статистика регрессии							
14		R^2	0,998427803907596						
15		Среднеквадратическое отклонение	0,664382718271549						
16		Число переменных X	2						
17		Наблюдения	19						
18		Скорректированный R^2	0,998231279396045						
19									
20		Дисперсионный анализ (ANOVA)							
21		df	SS	MS	F	Значимость F			
22		Регрессия	2	4485,04279281649	2242,52139641	5080,42379	3,7329631E-23		
23		Остатки	16	7,06247034140628	0,44140439634				
24		Итого	18	4492,10526315789					
25									
26		Уровень доверия	0,95						
27									
28		Коэффициенты	Среднеквадратическое отклонение	t-статистика	P-значение	Нижний 95%	Верхний 95%		
29		Пересечение	-2,28702739628397	0,865664125994555	-2,64193389515	0,01775891	-4,12215336433	-0,4519014282	
30		X1	1,66229790873236	0,129008125551406	12,8852186762	7,2881E-10	1,38881289973	1,9357829177	
31		X2	0,158855339068342	0,06470373521014	2,45511852681	0,02590594	0,02168954792	0,2960211302	
32									

Для нашего случая для каждого участка данное уравнение будет иметь вид

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2,$$

где y — изучаемый параметр, т. е. количество преступлений;

a_0 — число, стоящее на пересечении «Коэффициенты» и «Пересечение»;

a_i — числа, стоящие на пересечении «Коэффициенты» и «X»;

x_1 — количество осужденных;

x_2 — количество раскрытых преступлений.

Сделаем проверку, например, для десятого участка. В столбец «Проверка» в десятую строчку вводим вышеуказанную формулу.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений	Проверка	Прогноз					
3		1.	54	28	51			Регрессия				
4		2.	65	35	60			Регрессионная модель	Линейная			
5		3.	85	45	81			Необработанный вывод ЛИНЕЙН				
6		4.	68	36	62			0,158855339068342	1,66229790873236	-2,28702739628397		
7		5.	95	50	92			0,06470373521014	0,129008125551406	0,865664125994555		
8		6.	53	29	46			0,998427803907596	0,664382718271549	#Н/Д		
9		7.	67	36	62			5080,42379054967	16	#Н/Д		
10		8.	61	33	59			4485,04279281649	7,06247034140628	#Н/Д		
11		9.	59	32	53							
12		10.	74	39	73	=I29+I30*D12+I31*E12						
13		11.	58	31	50			Статистика регрессии				
14		12.	43	24	40			R^2	0,998427803907596			
15		13.	76	40	71			Среднеквадратическое отклонение	0,664382718271549			
16		14.	80	42	76			Число переменных X	2			
17		15.	72	38	70			Наблюдения	19			
18		16.	93	49	85			Скорректированный R^2	0,998231279396045			
19		17.	56	30	55			Дисперсионный анализ (ANOVA)				
20		18.	74	39	71							
21		19.	34	19	29							
22								df	SS	MS	F	
23								Регрессия	2	4485,04279281649	2242,52139641	5080,42379
24								Остатки	16	7,06247034140628	0,44140439634	
25								Итого	18	4492,10526315789		
26								Уровень доверия	0,95			
27												
28								Коэффициенты		Среднеквадратическое отклонение	t-статистика	P-значение
29								Пересечение	-2,28702739628397	0,865664125994555	-2,6419338952	0,01775891
30								X1	1,66229790873236	0,129008125551406	12,8852186762	7,2881E-10
31								X2	0,158855339068342	0,06470373521014	2,45511852681	0,02590594
32												

Нажимаем «Ввод» и получаем значение количества преступлений на нашем участке. Погрешность обусловлена точностью вычислений, достоверность которых мы выбрали 95%.

	A	B	C	D	E	F
1						
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений	Проверка
3		1.	54	28	51	
4		2.	65	35	60	
5		3.	85	45	81	
6		4.	68	36	62	
7		5.	95	50	92	
8		6.	53	29	46	
9		7.	67	36	62	
10		8.	61	33	59	
11		9.	59	32	53	
12		10.	74	39	73	74,1390308
13		11.	58	31	50	
14		12.	43	24	40	
15		13.	76	40	71	
16		14.	80	42	76	
17		15.	72	38	70	
18		16.	93	49	85	
19		17.	56	30	55	
20		18.	74	39	71	
21		19.	34	19	29	
22						

2. Сделаем прогноз: каким будет количество преступлений на 5 участке, если количество осужденных (переменная «X1») уменьшится на 10, а количество раскрытых преступлений (переменная «X2») увеличится на 15.

Для данного случая уравнение регрессии будет иметь вид

$$y = a_0 + a_1 \cdot (x1 - 10) + a_2 \cdot (x2 + 15),$$

где y — изучаемый параметр, т. е. количество преступлений;

a_0 — число, стоящее на пересечении «Коэффициенты» и «Пересечение»;

a_i — числа, стоящие на пересечении «Коэффициенты» и «X»;

$x1$ — количество осужденных;

$x2$ — количество раскрытых преступлений.

	A	B	C	D	E	F	G	H	I	J
1										
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений	Проверка	Прогноз			
3		1.	54	28	51			Регрессия		
4		2.	65	35	60			Регрессионная модель	Линейная	
5		3.	85	45	81					
6		4.	68	36	62			Необработанный вывод ЛИНЕЙН		
7		5.	95	50	92		=129+150*(D7-10)+131*(E7+15)	0,158855339068342	1,66229790873236	-2,28702739628397
8		6.	53	29	46			0,06470373521014	0,129008125551406	0,865664125994555
9		7.	67	36	62			0,998427803907596	0,664382718271549	#Н/Д
10		8.	61	33	59			5080,42379054967	16	#Н/Д
11		9.	59	32	53			4485,04279281649	7,06247034140628	#Н/Д
12		10.	74	39	73	74,139031				
13		11.	58	31	50			Статистика регрессии		
14		12.	43	24	40			R^2	0,998427803907596	
15		13.	76	40	71			Среднеквадратическое отклонение	0,664382718271549	
16		14.	80	42	76			Число переменных X	2	
17		15.	72	38	70			Наблюдения	19	
18		16.	93	49	85			Скорректированный R^2	0,998231279396045	
19		17.	56	30	55					
20		18.	74	39	71			Дисперсионный анализ (ANOVA)		
21		19.	34	19	29			df	SS	
22								Регрессия	2	4485,04279281649
23								Остатки	16	7,06247034140628
24								Итого	18	4492,10526315789
25								Уровень доверия	0,95	
26										
27								Коэффициенты	Среднеквадратическое отклонение	
28								Пересечение	-2,28702739628397	0,865664125994555
29								X1	1,66229790873236	0,129008125551406
30								X2	0,158855339068342	0,06470373521014
31										

Нажимаем «Ввод» и получаем значение количества преступлений на 5 участке, если количество осужденных уменьшится на 10, а количество раскрытых преступлений увеличится на 15.

	A	B	C	D	E	F	G
1							
2		№ п/п	Кол-во преступлений	Кол-во осужденных	Кол-во раскрытых преступлений	Проверка	Прогноз
3		1.	54	28	51		
4		2.	65	35	60		
5		3.	85	45	81		
6		4.	68	36	62		
7		5.	95	50	92		81,20241023
8		6.	53	29	46		
9		7.	67	36	62		
10		8.	61	33	59		
11		9.	59	32	53		
12		10.	74	39	73	74,139031	
13		11.	58	31	50		
14		12.	43	24	40		
15		13.	76	40	71		
16		14.	80	42	76		
17		15.	72	38	70		
18		16.	93	49	85		
19		17.	56	30	55		
20		18.	74	39	71		
21		19.	34	19	29		
22							

Как видно при данных условиях количество преступлений в данном районе составит 81, что на 14 правонарушений меньше, чем при исходных данных.

Пример 3. За 6 лет работы подразделения полиции были собраны ежеквартальные данные о количестве краж, совершенных на подконтрольной территории, значения которых сведены в таблицу.

Год	№ квартала	Количество краж
1	1	45
	2	55
	3	70
	4	50
2	5	51
	6	59
	7	75
	8	53
3	9	55
	10	63
	11	79
	12	62
4	13	60
	14	67
	15	85
	16	63
5	17	65
	18	77
	19	91
	20	69
6	21	70
	22	83
	23	94
	24	73

Необходимо: на основании имеющихся данных сделать прогноз оценки количества краж на год, следующий за отчетным периодом.

Решение.

Для решения данной задачи воспользуемся редактором Calc.

Все вычисления можно производить на одном листе. Для наглядности скриншоты выполнения программы будут отдельными элементами.

Заносим данные в виде таблицы.

Без имени 1.xls - LibreOffice Calc

Файл Правка Вид Вставка Форм

PT Astra Sans 12 Ж

R34 f_x ∑ =

	A	B	C
1	№ квартала	Количество краж	
2	1	45	
3	2	55	
4	3	70	
5	4	50	
6	5	51	
7	6	59	
8	7	75	
9	8	53	
10	9	55	
11	10	63	
12	11	79	
13	12	62	
14	13	60	
15	14	67	
16	15	85	
17	16	63	
18	17	65	
19	18	77	
20	19	91	
21	20	69	
22	21	70	
23	22	83	
24	23	94	
25	24	73	
26			

Описанный ниже алгоритм справедлив для данных, повторяющихся с определенной периодичностью. Очевидно, что количество краж из года в год в различные сезоны либо повышается, либо понижается, так летом их значение увеличивается в связи с сезоном отпусков, и т. п. Проверим, имеют ли наши данные цикличность. Для этого копируем данные о количестве краж и вставляем их в любом свободном месте листа, рядом вставляем со сдвигом на одну ячейку вниз и т. д. (чем больше, тем лучше).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1																											
2																											
3			45																								
4			55	45																							
5			70	55																							
6			50	70	55	45																					
7			55	50	70	55	45																				
8			59	51	50	70	55	45																			
9			75	59	51	50	70	55	45																		
10			53	75	59	51	50	70	55	45																	
11			55	53	75	59	51	50	70	55	45																
12			63	55	53	75	59	51	50	70	55	45															
13			79	63	55	53	75	59	51	50	70	55	45														
14			62	79	63	55	53	75	59	51	50	70	55	45													
15			60	62	79	63	55	53	75	59	51	50	70	55	45												
16			67	60	62	79	63	55	53	75	59	51	50	70	55	45											
17			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45										
18			63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45									
19			65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45								
20			77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45							
21			91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
22			69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45					
23			70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45				
24			83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45			
25			84	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45		
26			73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
27				73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
28					73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55
29						73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70
30							73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50
31								73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51
32									73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59
33										73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75
34											73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53
35												73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55
36													73	94	83	70	69	91	77	65	63	85	67	60	62	79	63
37														73	94	83	70	69	91	77	65	63	85	67	60	62	79
38															73	94	83	70	69	91	77	65	63	85	67	60	62
39																73	94	83	70	69	91	77	65	63	85	67	60
40																	73	94	83	70	69	91	77	65	63	85	67
41																		73	94	83	70	69	91	77	65	63	85
42																			73	94	83	70	69	91	77	65	63
43																				73	94	83	70	69	91	77	65
44																					73	94	83	70	69	91	77
45																						73	94	83	70	69	91
46																							73	94	83	70	69
47																								73	94	83	70
48																									73	94	83
49																										73	94
50																											73
51																											

«Обрезаем» полученную таблицу по первому столбцу.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1																											
2																											
3			45																								
4			55	45																							
5			70	55	45																						
6			50	70	55	40																					
7			51	50	70	55	45																				
8			59	51	50	70	55	45																			
9			75	59	51	50	70	55	45																		
10			53	75	59	51	50	70	55	45																	
11			55	53	75	59	51	50	70	55	45																
12			63	55	53	75	59	51	50	70	55	45															
13			79	63	55	53	75	59	51	50	70	55	45														
14			62	79	63	55	53	75	59	51	50	70	55	45													
15			60	62	79	63	55	53	75	59	51	50	70	55	45												
16			67	60	62	79	63	55	53	75	59	51	50	70	55	45											
17			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45										
18			63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45									
19			65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45								
20			77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45							
21			91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
22			69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45					
23			70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45				
24			83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45			
25			94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45		
26			73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
27				73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
28				94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
29				73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
30					73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55
31						73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70
32							73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50
33								73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51
34									73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59
35										73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75
36											73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	75
37												73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	
38													73	94	83	70	69	91	77	65	63	85	67	60	62	79	
39														73	94	83	70	69	91	77	65	63	85	67	60	62	
40															73	94	83	70	69	91	77	65	63	85	67	60	
41																73	94	83	70	69	91	77	65	63	85	67	
42																	73	94	83	70	69	91	77	65	63	85	
43																		73	94	83	70	69	91	77	65	63	
44																			73	94	83	70	69	91	77	65	
45																				73	94	83	70	69	91	77	
46																					73	94	83	70	69	91	
47																						73	94	83	70	69	
48																							73	94	83	70	
49																								73	94	83	
50																									73	94	
51																										73	

В поле «Входной диапазон:» вводим номера ячеек, получившейся треугольной таблицы.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1																							
2																							
3																							
4			45																				
5			55	45																			
6			70	55	45																		
7			50	70	55	45																	
8			51	50	70	55	45																
9			59	51	50	70	55	45															
10			75	59	51	50	70	55	45														
11			53	75	59	51	50	70	55	45													
12			63	55	53	75	59	51	50	70	55	45											
13			79	63	55	53	75	59	51	50	70	55	45										
14			62	79	63	55	53	75	59	51	50	70	55	45									
15			60	62	79	63	55	53	75	59	51	50	70	55	45								
16			67	60	62	79	63	55	53	75	59	51	50	70	55	45							
17			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
18			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
19			65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45				
20			77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45			
21			91	77	65	60	62	79	63	55	53	75	59	51	50	70	55	45					
22			69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
23			70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
24			83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55
25			94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70
26			73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50
27																							
28																							
29																							
30																							
31																							
32																							
33																							
34																							
35																							
36																							
37																							
38																							
39																							
40																							
41																							
42																							
43																							

Корреляция

Данные

Входной диапазон: \$Лист2.\$B\$3:\$W\$26

Результат в:

Группировать по

☒ Столбцам ☐ Строкам

Справка Отменить ОК

Выведем расчетные значения на этот же лист, выбрав «Результат в:» и указав ячейку (это левый верхний угол таблицы с результатами).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1																							
2																							
3																							
4			45																				
5			55	45																			
6			70	55	45																		
7			50	70	55	45																	
8			51	50	70	55	45																
9			59	51	50	70	55	45															
10			75	59	51	50	70	55	45														
11			53	75	59	51	50	70	55	45													
12			63	55	53	75	59	51	50	70	55	45											
13			79	63	55	53	75	59	51	50	70	55	45										
14			62	79	63	55	53	75	59	51	50	70	55	45									
15			60	62	79	63	55	53	75	59	51	50	70	55	45								
16			67	60	62	79	63	55	53	75	59	51	50	70	55	45							
17			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
18			85	67	60	62	79	63	55	53	75	59	51	50	70	55	45						
19			65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45				
20			77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45			
21			91	77	65	60	62	79	63	55	53	75	59	51	50	70	55	45					
22			69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
23			70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
24			83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55
25			94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70
26			73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50
27																							
28																							
29																							
30																							
31																							
32																							
33																							
34																							
35																							
36																							
37																							
38																							
39																							
40																							
41																							
42																							
43																							

Корреляция

Данные

Входной диапазон: \$Лист2.\$B\$3:\$W\$26

Результат в: \$B\$28

Группировать по

☒ Столбцам ☐ Строкам

Справка Отменить ОК

Нажимаем «ОК».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
22		69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45		
23		70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45	
24		83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55	45
25		94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70	55
26		73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51	50	70
27																							
28		Корреляции	Столбец 1	Столбец 2	Столбец 3	Столбец 4	Столбец 5	Столбец 6	Столбец 7	Столбец 8	Столбец 9	Столбец 10	Столбец 11	Столбец 12	Столбец 13	Столбец 14	Столбец 15	Столбец 16	Столбец 17	Столбец 18	Столбец 19	Столбец 20	Столбец 21
29		Столбец 1	1																				
30		Столбец 2	0,38215034	1																			
31		Столбец 3	0,0194752	0,38616762	1																		
32		Столбец 4	0,39456033	-0,0044342	0,28713262	1																	
33		Столбец 5	0,98665523	0,39011839	-0,0554095	0,27150414	1																
34		Столбец 6	0,25443619	0,98678599	0,39762038	-0,0867497	0,26632999	1															
35		Столбец 7	-0,1451138	0,26874773	0,98436306	0,28519974	-0,1467803	0,2703459	1														
36		Столбец 8	0,32243512	-0,1651092	0,13814095	0,98257652	0,27514168	-0,1745314	0,12409152	1													
37		Столбец 9	0,98023188	0,32070753	-0,2267986	0,12296313	0,98251688	0,2720511	-0,250944	0,10813871	1												
38		Столбец 10	0,10041957	0,98024642	0,35470957	-0,247739	0,12208747	0,98268223	0,30901884	-0,2729419	0,10716387	1											
39		Столбец 11	-0,3540253	0,12796393	0,97516595	0,21566648	-0,31748653	0,13376389	0,97481593	0,1579465	-0,3514526	0,13165769	1										
40		Столбец 12	0,2409431	-0,3552178	0,0301726	0,98395379	0,21225987	-0,3245657	0,0257227	0,98301417	0,15212458	-0,3546187	0,02129582	1									
41		Столбец 13	0,9767987	0,2411409	-0,205767	0,0295541	0,9839546	0,21218085	-0,4012301	0,02482489	0,98307668	0,15217575	-0,4408328	0,02037636	1								
42		Столбец 14	-0,04656	0,98357263	0,24342779	-0,4686427	0,0316103	0,98770299	0,21458097	-0,4468296	0,02599592	0,98884414	0,14006712	-0,4599193	0,02271468	1							
43		Столбец 15	-0,5858371	0,0123129	0,9785655	0,04928517	-0,5415617	0,08581891	0,9815076	0,00816913	-0,5266762	0,09352	0,98352756	0,0267305	-0,5312356	0,00575131	1						
44		Столбец 16	0,20498865	-0,5762443	-0,122102	0,98718993	0,05674615	-0,5361619	-0,057645	0,99189029	0,00981638	-0,518485	-0,0476235	0,98326858	0,03257218	-0,5447782	-0,1527813	1					
45		Столбец 17	0,9860729	0,1985316	-0,6390548	-0,1028985	0,98716626	0,0517033	-0,6309357	-0,033749	0,99197937	0,0037039	-0,6073353	-0,0369573	0,98317575	0,03616164	-0,6268096	-0,1419761	1				
46		Столбец 18	-0,2606557	0,9857486	0,3782791	-0,6263773	-0,1284247	0,98834423	0,2307456	-0,6182081	-0,0574159	0,99192069	0,17196609	-0,5997828	-0,0622695	0,99624846	0,2039083	-0,6170224	-0,1720841	1			
47		Столбец 19	-0,7523841	-0,0891782	0,98465363	0,16484994	-0,7038135	0,0118836	0,97810232	-0,0800811	-0,6987071	0,16072496	0,98629007	0,02981718	-0,6710367	-0,0128134	0,9930915	0,0442667	-0,6904283	0,0199643		1	
48		Столбец 20	0,25765711	-0,7370822	-0,2989621	0,99389268	0,19799907	-0,6859853	-0,2039929	0,99150164	-0,0578541	-0,6791853	-0,0110586	0,98800615	0,05538721	-0,6662828	-0,2346862	0,9929806	0,0721388	-0,6701631	-0,2043414	1	
49		Столбец 21	0,9801295	0,2211713	-0,7986546	-0,258572	0,99510556	0,16733201	-0,7732823	-0,1522862	0,99138387	-0,1084055	-0,8145547	0	0,98865748	0,05222626	-0,7418936	-0,2163856	0,9929528	0,0267011	-0,7668797	-0,1833843	1
50		Столбец 22	-0,5737472	0,9867398	0,9412189	-0,7783153	-0,4638914	0,9975104	0,9621032	-0,7512436	-0,2825625	0,9910607	0,7714543	-0,8563535	-0,1666185	0,9971765	0,976221	-0,7512436	-0,3726982	0,9971765	0,9532134	-0,7757194	-0,3499566
51																							

Полученная таблица дает результат сравнения каждого столбца с каждым на предмет схожести. Рассмотрим значения столбца «Столбец 1». В каждой строчке первый столбец нашей таблицы сравнивался с самим собой и с последующими. Максимальное значение при сравнении — 1, если столбцы не похожи, то значение стремиться к нулю. Очевидно, что первый столбец при сравнении с самим собой даст значение коэффициента сравнения 1. Выделим цветом те ячейки, в которых значение коэффициента сравнения приближается к 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
22		69	91	77	65	63	85	67	60	62	79	63	55	53	75	59	51
23		70	69	91	77	65	63	85	67	60	62	79	63	55	53	75	59
24		83	70	69	91	77	65	63	85	67	60	62	79	63	55	53	75
25		94	83	70	69	91	77	65	63	85	67	60	62	79	63	55	53
26		73	94	83	70	69	91	77	65	63	85	67	60	62	79	63	55
27																	
28		Корреляции	Столбец 1	Столбец 2	Столбец 3	Столбец 4	Столбец 5	Столбец 6	Столбец 7	Столбец 8	Столбец 9	Столбец 10	Столбец 11	Столбец 12	Столбец 13	Столбец 14	Столбец 15
29		Столбец 1	1														
30		Столбец 2	0,38215034	1													
31		Столбец 3	0,01947525	0,38616762	1												
32		Столбец 4	0,39456033	-0,0044342	0,28713262	1											
33		Столбец 5	0,98665523	0,39011839	-0,0554095	0,27150414	1										
34		Столбец 6	0,25443619	0,98678599	0,39762038	-0,0867497	0,26632999	1									
35		Столбец 7	-0,1451138	0,26874773	0,98436306	0,28519974	-0,1467803	0,2703459	1								
36		Столбец 8	0,32243512	-0,1651092	0,13814095	0,98257652	0,27514168	-0,1745314	0,12409152	1							
37		Столбец 9	0,98023188	0,32070753	-0,2267986	0,12296313	0,98251688	0,2720511	-0,250944	0,10813871	1						
38		Столбец 10	0,10041957	0,98024642	0,35470957	-0,247739	0,12208747	0,98268223	0,30901884	-0,2729419	0,10716387	1					
39		Столбец 11	-0,3540253	0,12796393	0,97516595	0,21566648	-0,31748653	0,13376389	0,97481593	0,1579465	-0,3514526	0,13165769	1				
40		Столбец 12	0,2409431	-0,3552178	0,0301726	0,98395379	0,21225987	-0,3245657	0,0257227	0,98301417	0,15212458	-0,3546187	0,02129582	1			
41		Столбец 13	0,9767987	0,24114088	-0,205767	0,02955411	0,9839546	0,21218085	-0,4012301	0,02482489	0,98307668	0,15217575	-0,4408328	0,02037636	1		
42		Столбец 14	-0,04656	0,98357263	0,24342779	-0,4686427	0,0316103	0,98770299	0,21458097	-0,4468296	0,02599592	0,98884414	0,14006712	-0,4599193	0,02271468	1	
43		Столбец 15	-0,5858371	0,0123129	0,97856554	0,04928517	-0,5415617	0,08581891	0,9815076	0,00816913	-0,5266762	0,09352	0,98352756	0,0267305	-0,5312356	0,00575131	1
44		Столбец 16	0,20498865	-0,5762443	-0,122102	0,98718993	0,05674615	-0,5361619	-0,057645	0,99189029	0,00981638	-0,518485	-0,0476235	0,98326858	0,03257218	-0,5447782	-0,1527813
45		Столбец 17	0,9860729	0,19853165	-0,6390548	-0,1028985	0,98716626	0,05170326	-0,6309357	-0,033749	0,99197937	0,00370387	-0,6073353	-0,0369573	0,98317575	0,03616164	-0,6268096
46		Столбец 18	-0,2606557	0,98574857	0,37827907	-0,6263773	-0,1284247	0,98834423	0,2307456	-0,6182081	-0,0574159	0,99192069	0,17196609	-0,5997828	-0,0622695	0,99624846	0,20390826
47		Столбец 19	-0,7523841	-0,0891782	0,98465363	0,16484994	-0,7038135	0,0118836	0,97810232	-0,0800811	-0,6987071	0,16072496	0,98629007	0,02981718	-0,6710367	-0,0128134	0,99309145
48		Столбец 20	0,25765711	-0,7370822	-0,2989621	0,99389268	0,19799907	-0,6859853	-0,2039929	0,99150164	-0,0578541	-0,6791853	-0,0110586	0,98800615	0,05538721	-0,6662828	-0,2348682
49		Столбец 21	0,98012946	0,22117125	-0,7986546	-0,258572	0,99510556	0,16733201	-0,7732823	-0,1522862	0,99138387	-0,1084055	-0,8145547	0	0,98865748	0,05222626	-0,7418936
50		Столбец 22	-0,5737472	0,98673898	0,9412189	-0,7783153	-0,4638914	0,99751037	0,96210321	-0,7512436	-0,2825625	0,99106067	0,77145428	-0,8563535	-0,1666185	0,99717646	0,97622104
51																	

Видно, что ячейки выделенные красным цветом, повторяются с периодичностью 4, то есть у такого преступления как кражи есть сезонность, или мы проверили, что в году четыре квартала.

Теперь перейдем к дальнейшим вычислениям. Рассчитаем скользящую среднюю. Обратите внимание, что значения 1 и 5 квартала поделены на 2.

Файл Правка Вид Вставка Формат Стили Лист Данные					
PT Astra Sans 11 Ж К Ч Т					
СТОТКЛ.Г f _x $\times$ $\checkmark$ $= (B2/2 + B3 + B4 + B5 + B6/2) / 4$					
	A	B	C	D	
1	№ квартала	Количество краж	Скользящая средняя		
2	1	45			
3	2	55			
4	3	70	$= (B2/2 + B3 + B4 + B5 + B6/2) / 4$		
5	4	50			
6	5	51			
7	6	59			
8	7	75			
9	8	53			
10	9	55			
11	10	63			
12	11	79			
13	12	62			
14	13	60			
15	14	67			
16	15	85			
17	16	63			
18	17	65			
19	18	77			
20	19	91			
21	20	69			
22	21	70			
23	22	83			
24	23	94			
25	24	73			

	A	B	C	D	
1	№ квартала	Количество краж	Скользящая средняя		
2	1	45			
3	2	55			
4	3	70	55,75		
5	4	50			
6	5	51			
7	6	59			
8	7	75			
9	8	53			
10	9	55			
11	10	63			
12	11	79			
13	12	62			
14	13	60			
15	14	67			
16	15	85			
17	16	63			
18	17	65			
19	18	77			
20	19	91			
21	20	69			
22	21	70			
23	22	83			
24	23	94			
25	24	73			

И копируем на весь диапазон данных. Расчетные значения, по очевидным причинам, в первые и в последние две строчки не вводятся (при расчете в 1, 2, 23 и 24 квартале вычисления по данной формуле не дадут правдивого результата, значения данных ячеек определим ниже).

Подведем курсор к правому нижнему краю выделенной ячейки и скопируем вычисления на весь диапазон.

	A	B	C	D	
1	№ квартала	Количество краж	Скользящая средняя		
2	1	45			
3	2	55			
4	3	70	55,75		
5	4	50	57		
6	5	51	58,125		
7	6	59	59,125		
8	7	75	60		
9	8	53	61		
10	9	55	62		
11	10	63	63,625		
12	11	79	65,375		
13	12	62	66,5		
14	13	60	67,75		
15	14	67	68,625		
16	15	85	69,375		
17	16	63	71,25		
18	17	65	73,25		
19	18	77	74,75		
20	19	91	76,125		
21	20	69	77,5		
22	21	70	78,625		
23	22	83	79,5		
24	23	94			
25	24	73			
26					

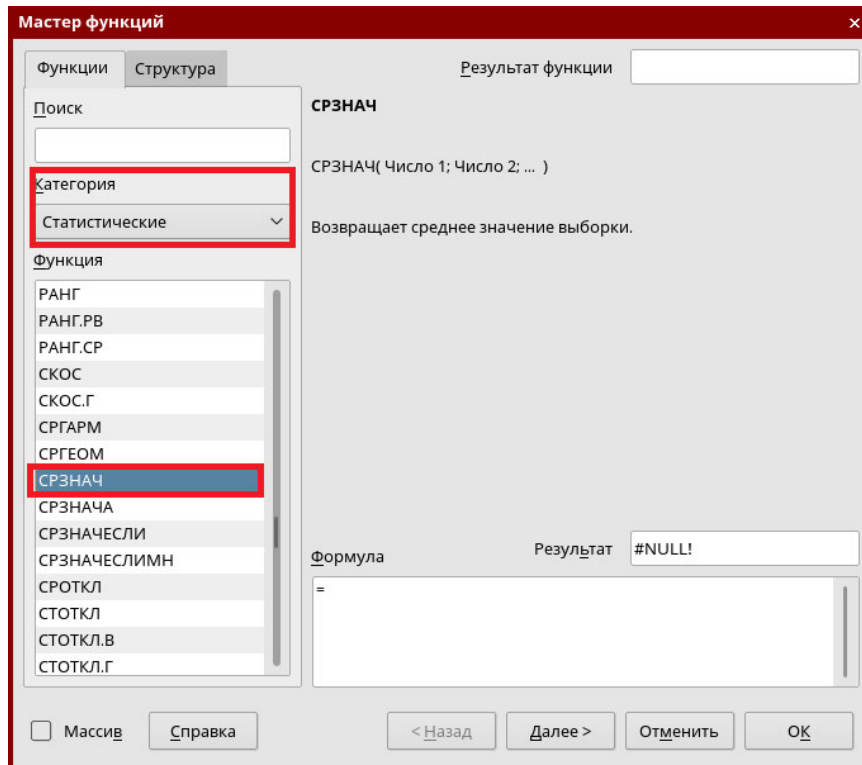
Произведем оценку сезонной компоненты.

	A	B	C	D	
1	№ квартала	Количество кraj	Скользящая средняя	Оценка сезонной компоненты	
2	1	45			
3	2	55			
4	3	70	55,75	=B4-C4	
5	4	50	57		
6	5	51	58,125		
7	6	59	59,125		
8	7	75	60		
9	8	53	61		
10	9	55	62		
11	10	63	63,625		
12	11	79	65,375		
13	12	62	66,5		
14	13	60	67,75		
15	14	67	68,625		
16	15	85	69,375		
17	16	63	71,25		
18	17	65	73,25		
19	18	77	74,75		
20	19	91	76,125		
21	20	69	77,5		
22	21	70	78,625		
23	22	83	79,5		
24	23	94			
25	24	73			
26					

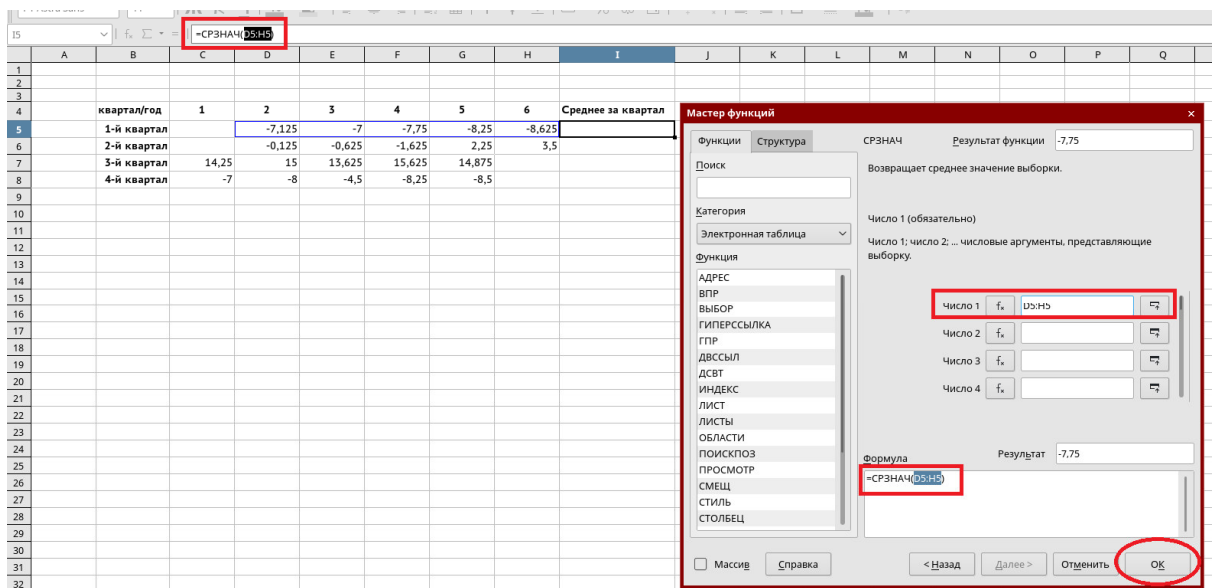
Продолжим вычисления на весь диапазон.

	A	B	C	D	E
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	
2	1	45			
3	2	55			
4	3	70	55,75	14,25	
5	4	50	57	-7	
6	5	51	58,125	-7,125	
7	6	59	59,125	-0,125	
8	7	75	60	15	
9	8	53	61	-8	
10	9	55	62	-7	
11	10	63	63,625	-0,625	
12	11	79	65,375	13,625	
13	12	62	66,5	-4,5	
14	13	60	67,75	-7,75	
15	14	67	68,625	-1,625	
16	15	85	69,375	15,625	
17	16	63	71,25	-8,25	
18	17	65	73,25	-8,25	
19	18	77	74,75	2,25	
20	19	91	76,125	14,875	
21	20	69	77,5	-8,5	
22	21	70	78,625	-8,625	
23	22	83	79,5	3,5	
24	23	94			
25	24	73			

Далее вычислим среднее по каждому кварталу за 6 лет, подгружаем «Мастер функций»: «Вставка — Функция» или воспользуемся пиктограммой «Мастер функций». Далее в поле «Категория» выбираем «Статистические». В поле «Функция» находим «СРЗНАЧ».



Нажимаем «Далее». В поле «Число 1» вводим диапазон ячеек с нашими данными. Нажимаем «ОК».



Значения первого и второго кварталов для первого года в вычислениях не учитываются, так же как и значения третьего и четвертого для 6 года.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3																	
4		квартал/год	1	2	3	4	5	6	Среднее за квартал								
5		1-й квартал		-7,125	-7	-7,75	-8,25	-8,625	-7,75								
6		2-й квартал		-0,125	-0,625	-1,625	2,25	3,5	0,675								
7		3-й квартал		14,25	15	13,625	15,625	14,875									
8		4-й квартал		-7	-8	-4,5	-8,25	-8,5									
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	

=СРЗНАЧ(C7:G7)

Мастер функций

Функции
Структура

Поиск

Категория Электронная таблица

Функция

АДРЕС
ВПР
ВЫБОР
ГИПЕРССЫЛКА
ГПР
ДВССЫЛ
ДСВТ
ИНДЕКС
ЛИСТ
ЛИСТЫ
ОБЛАСТИ
ПОИСКПОЗ
ПРОСМОТР
СМЕЩ
СТИЛЬ
СТОЛБЕЦ

Число 1 (обязательно)
Число 1; число 2 ... ключевые аргументы, представляющие выборку.

Число 1 f_x C7:G7

Число 2 f_x

Число 3 f_x

Число 4 f_x

Результат функции 14,675

Число 1 (обязательно)
Число 1; число 2 ... ключевые аргументы, представляющие выборку.

Формула

=СРЗНАЧ(C7:G7)

Результат 14,675

Массив
Справка

< Назад
Далее >
Отменить
ОК

Согласно третьего закона Ньютона если где-то прибавляется, то где-то должно уменьшиться, тогда сумма значений «Среднее за квартал» должно равняться нулю.

[illegible][illegible]

Так как сумма не равна нулю, то найдем поправочный коэффициент как среднее от среднего за квартал.

[illegible]

Получаем.

[illegible]

Далее найдем исправленное среднее.

[illegible]

И посчитаем данное значение для каждого квартала. Значение буквы ячейки с поправкой ограничим знаком \$, для закрепления значений данной ячейки. Оставшиеся значения исправленной средней найдем путем подведения курсора к правому нижнему краю ячейки и перетаскиванием на нужный диапазон.

[illegible]

Проверим правильность наших вычислений и справедливость третьего закона Ньютона, т. е. найдем сумму исправленного среднего, которая должна равняться нулю.

[illegible]

Получили ноль.

[illegible]

Значения исправленного среднего по каждому кварталу вносим в столбец сезонной компоненты общих вычислений (предлагается после копирования значений исправленного среднего использовать специальную вставку как значение). В результате получим.

	A	B	C	D	E	F
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	
2	1	45			-7,8375	
3	2	55			0,5875	
4	3	70	55,75	14,25	14,5875	
5	4	50	57	-7	-7,3375	
6	5	51	58,125	-7,125	-7,8375	
7	6	59	59,125	-0,125	0,5875	
8	7	75	60	15	14,5875	
9	8	53	61	-8	-7,3375	
10	9	55	62	-7	-7,8375	
11	10	63	63,625	-0,625	0,5875	
12	11	79	65,375	13,625	14,5875	
13	12	62	66,5	-4,5	-7,3375	
14	13	60	67,75	-7,75	-7,8375	
15	14	67	68,625	-1,625	0,5875	
16	15	85	69,375	15,625	14,5875	
17	16	63	71,25	-8,25	-7,3375	
18	17	65	73,25	-8,25	-7,8375	
19	18	77	74,75	2,25	0,5875	
20	19	91	76,125	14,875	14,5875	
21	20	69	77,5	-8,5	-7,3375	
22	21	70	78,625	-8,625	-7,8375	
23	22	83	79,5	3,5	0,5875	
24	23	94			14,5875	
25	24	73			-7,3375	
26						

Проведем оценку тренда.

	A	B	C	D	E	F	G
1	№ квартала	Количество кв.ж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	
2	1	45			-7,8375	=B2-E2	
3	2	55			0,5875		
4	3	70	55,75	14,25	14,5875		
5	4	50	57	-7	-7,3375		
6	5	51	58,125	-7,125	-7,8375		
7	6	59	59,125	-0,125	0,5875		
8	7	75	60	15	14,5875		
9	8	53	61	-8	-7,3375		
10	9	55	62	-7	-7,8375		
11	10	63	63,625	-0,625	0,5875		
12	11	79	65,375	13,625	14,5875		
13	12	62	66,5	-4,5	-7,3375		
14	13	60	67,75	-7,75	-7,8375		
15	14	67	68,625	-1,625	0,5875		
16	15	85	69,375	15,625	14,5875		
17	16	63	71,25	-8,25	-7,3375		
18	17	65	73,25	-8,25	-7,8375		
19	18	77	74,75	2,25	0,5875		
20	19	91	76,125	14,875	14,5875		
21	20	69	77,5	-8,5	-7,3375		
22	21	70	78,625	-8,625	-7,8375		
23	22	83	79,5	3,5	0,5875		
24	23	94			14,5875		
25	24	73			-7,3375		
26							
27							

Вычислим данное значение для каждого квартала.

	A	B	C	D	E	F	G
1	№ квартала	Количество кварт	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	
2	1	45			-7,8375	52,8375	
3	2	55			0,5875	54,4125	
4	3	70	55,75	14,25	14,5875	55,4125	
5	4	50	57	-7	-7,3375	57,3375	
6	5	51	58,125	-7,125	-7,8375	58,8375	
7	6	59	59,125	-0,125	0,5875	58,4125	
8	7	75	60	15	14,5875	60,4125	
9	8	53	61	-8	-7,3375	60,3375	
10	9	55	62	-7	-7,8375	62,8375	
11	10	63	63,625	-0,625	0,5875	62,4125	
12	11	79	65,375	13,625	14,5875	64,4125	
13	12	62	66,5	-4,5	-7,3375	69,3375	
14	13	60	67,75	-7,75	-7,8375	67,8375	
15	14	67	68,625	-1,625	0,5875	66,4125	
16	15	85	69,375	15,625	14,5875	70,4125	
17	16	63	71,25	-8,25	-7,3375	70,3375	
18	17	65	73,25	-8,25	-7,8375	72,8375	
19	18	77	74,75	2,25	0,5875	76,4125	
20	19	91	76,125	14,875	14,5875	76,4125	
21	20	69	77,5	-8,5	-7,3375	76,3375	
22	21	70	78,625	-8,625	-7,8375	77,8375	
23	22	83	79,5	3,5	0,5875	82,4125	
24	23	94			14,5875	79,4125	
25	24	73			-7,3375	80,3375	
26							

Определим тренд. Для этого оценим регрессию между оценкой тренда и № квартала.

В меню «Данные» выбираем «Статистика» → «Регрессия...».

В поле «Диапазон независимых переменных (X)» вводим значения ячеек «№ квартала».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда								
2	1	45			-7,8375	52,8375								
3	2	55			0,5875	54,4125								
4	3	70	55,75	14,25	14,5875	55,4125								
5	4	50	57	-7	-7,3375	57,3375								
6	5	51	58,125	-7,125	-7,8375	58,8375								
7	6	59	59,125	-0,125	0,5875	58,4125								
8	7	75	60	15	14,5875	60,4125								
9	8	53	61	-8	-7,3375	60,3375								
10	9	55	62	-7	-7,8375	62,8375								
11	10	63	63,625	-0,625	0,5875	62,4125								
12	11	79	65,375	13,625	14,5875	64,4125								
13	12	62	66,5	-4,5	-7,3375	69,3375								
14	13	60	67,75	-7,75	-7,8375	67,8375								
15	14	67	68,625	-1,625	0,5875	66,4125								
16	15	85	69,375	15,625	14,5875	70,4125								
17	16	63	71,25	-8,25	-7,3375	70,3375								
18	17	65	73,25	-8,25	-7,8375	72,8375								
19	18	77	74,75	2,25	0,5875	76,4125								
20	19	91	76,125	14,875	14,5875	76,4125								
21	20	69	77,5	-8,5	-7,3375	76,3375								
22	21	70	78,625	-8,625	-7,8375	77,8375								
23	22	83	79,5	3,5	0,5875	82,4125								
24	23	94			14,5875	79,4125								
25	24	73			-7,3375	80,3375								
26														
27														
28														

Регрессия

Данные

Диапазон независимых переменных (X): \$Лист1.\$A\$2:\$A\$25

Диапазон зависимой переменной (Y):

☐ Оба диапазона X и Y с подписями

Результат в:

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия: 0,95

☒ Вычислить остатки ☐ Свободный член равен 0

Неверный диапазон зависимых переменных.

Справка Отменить ОК

В поле «Диапазон зависимой переменной (Y)» вводим номера ячеек, содержащие оценку тренда. В поле «Результат в:» выбираем удобную для размещения результатов расчета ячейку. В данном случае достоверность полученных данных по умолчанию оценивается 95 % (0,95). Тогда вероятность ошибки — 5% (0,05). Это значение можно изменить, выбрав в окне «Регрессия» «Уровень доверия», и установить нужное значение.

Поставим галочку в поле «Вычислить остатки».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда								
2	1	45			-7,8375	52,8375								
3	2	55			0,5875	54,4125								
4	3	70	55,75	14,25	14,5875	55,4125								
5	4	50	57	-7	-7,3375	57,3375								
6	5	51	58,125	-7,125	-7,8375	58,8375								
7	6	59	59,125	-0,125	0,5875	58,4125								
8	7	75	60	15	14,5875	60,4125								
9	8	53	61	-8	-7,3375	60,3375								
10	9	55	62	-7	-7,8375	62,8375								
11	10	63	63,625	-0,625	0,5875	62,4125								
12	11	79	65,375	13,625	14,5875	64,4125								
13	12	62	66,5	-4,5	-7,3375	69,3375								
14	13	60	67,75	-7,75	-7,8375	67,8375								
15	14	67	68,625	-1,625	0,5875	66,4125								
16	15	85	69,375	15,625	14,5875	70,4125								
17	16	63	71,25	-8,25	-7,3375	70,3375								
18	17	65	73,25	-8,25	-7,8375	72,8375								
19	18	77	74,75	2,25	0,5875	76,4125								
20	19	91	76,125	14,875	14,5875	76,4125								
21	20	69	77,5	-8,5	-7,3375	76,3375								
22	21	70	78,625	-8,625	-7,8375	77,8375								
23	22	83	79,5	3,5	0,5875	82,4125								
24	23	94			14,5875	79,4125								
25	24	73			-7,3375	80,3375								
26														
27														

Регрессия

Данные

Диапазон независимых переменных (X): \$Лист1.\$A\$2:\$A\$25

Диапазон зависимой переменной (Y): \$Лист1.\$F\$2:\$F\$24

☐ Оба диапазона X и Y с подписями

Результат в: \$W\$29

Группировать по

☒ Столбцы ☐ Строки

Варианты вывода регрессии

☒ Линейная регрессия

☐ Логарифмическая регрессия

☐ Степенная регрессия

Параметры

Уровень доверия: 0,95

☒ Вычислить остатки ☐ Свободный член равен 0

Одномерная регрессия: число наблюдений X и Y должно совпадать.

Справка Отменить ОК

В результате получаем.

O	P	Q	R	S	T	U	V
	1829,52195652174	43,56179347826					
Статистика регрессии							
R ²	0,976743275105						
Среднеквадратическое отклонение	1,407153695138						
Число переменных X	1						
Наблюдения	24						
Скорректированный R ²	0,975686151247						
Дисперсионный анализ (ANOVA)							
	df	SS	MS	F	Значимость F		
Регрессия	1	1829,52195652174	1829,5219565	923,962946	1,8302451E-19		
Остатки	22	43,5617934782608	1,9800815217				
Итог	23	1873,08375					
Уровень доверия	0,95						
	Коэффициенты	Среднеквадратическое отклонение	t-статистика	P-значение	Нижний 95%	Верхний 95%	
Пересечение	51,48369565217	0,592904893041722	86,832974827	2,1283E-29	50,2540861627	52,713305142	
X1	1,261304347826	0,041494698684333	30,396758811	1,8302E-19	1,17524960976	1,3473590859	
X1	Прогнозное Y	Y	Остатки				
	1	52,745	52,8375	0,0925			
	2	54,00630434783	54,4125	0,4061956522			
	3	55,26760869565	55,4125	0,1448913043			
	4	56,52891304348	57,3375	0,8085869565			
	5	57,7902173913	58,8375	1,0472826087			
	6	59,05152173913	58,4125	-0,639021739			

Проверяем полученные результаты на правильность вычислений. Значение множественного коэффициента регрессии должно быть близким к 1 («R²» — выделено красным цветом). В противном случае, все имеющиеся входные статистические данные сфальсифицированы и не имеют ничего общего с реальной обстановкой.

Значение «Значимость F» (выделено красным цветом) дисперсионного анализа должно быть меньше 0,05, что было определено выше, и указывает на величину ошибки.

Оба этих параметра в нашем примере соответствуют указанным диапазонам, поэтому полученным результатам можно доверять с вероятностью 95%.

«P-Значение» переменной «Переменная X1» — меньше 0,05. Таким образом, полученным результатам можно доверять.

Значения «Прогнозное Y» является трендом, а «Остатки» — случайной компонентой.

Копируем значения «Прогнозное Y».

О	Р	Q	R	S	T
	Пересечение	51,48369565217	0,592904893041722	86,832974827	2,1283E-29
	X1	1,261304347826	0,041494698684333	30,396758811	1,8302E-19
	X1	Прогнозное Y	Y	Остатки	
	1	52,745	52,8375	0,0925	
	2	54,00630434783	54,4125	0,4061956522	
	3	55,26760869565	55,4125	0,1448913043	
	4	56,52891304348	57,3375	0,8085869565	
	5	57,7902173913	58,8375	1,0472826087	
	6	59,05152173913	58,4125	-0,639021739	
	7	60,31282608696	60,4125	0,099673913	
	8	61,57413043478	60,3375	-1,236630435	
	9	62,83543478261	62,8375	0,0020652174	
	10	64,09673913043	62,4125	-1,68423913	
	11	65,35804347826	64,4125	-0,945543478	
	12	66,61934782609	69,3375	2,7181521739	
	13	67,88065217391	67,8375	-0,043152174	
	14	69,14195652174	66,4125	-2,729456522	
	15	70,40326086957	70,4125	0,0092391304	
	16	71,66456521739	70,3375	-1,327065217	
	17	72,92586956522	72,8375	-0,088369565	
	18	74,18717391304	76,4125	2,225326087	
	19	75,44847826087	76,4125	0,9640217391	
	20	76,7097826087	76,3375	-0,372282609	
	21	77,97108695652	77,8375	-0,133586957	
	22	79,23239130435	82,4125	3,1801086957	
	23	80,49369565217	79,4125	-1,081195652	
	24	81,755	80,3375	-1,4175	

Вставляем в исходную таблицу.

	A	B	C	D	E	F	G	H
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	
2	1	45			-7,8375	52,8375	52,745	
3	2	55			0,5875	54,4125	54,0063043	
4	3	70	55,75	14,25	14,5875	55,4125	55,2676087	
5	4	50	57	-7	-7,3375	57,3375	56,528913	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,7902174	
7	6	59	59,125	-0,125	0,5875	58,4125	59,0515217	
8	7	75	60	15	14,5875	60,4125	60,3128261	
9	8	53	61	-8	-7,3375	60,3375	61,5741304	
10	9	55	62	-7	-7,8375	62,8375	62,8354348	
11	10	63	63,625	-0,625	0,5875	62,4125	64,0967391	
12	11	79	65,375	13,625	14,5875	64,4125	65,3580435	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,6193478	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,8806522	
15	14	67	68,625	-1,625	0,5875	66,4125	69,1419565	
16	15	85	69,375	15,625	14,5875	70,4125	70,4032609	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,6645652	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,9258696	
19	18	77	74,75	2,25	0,5875	76,4125	74,1871739	
20	19	91	76,125	14,875	14,5875	76,4125	75,4484783	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,7097826	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971087	
23	22	83	79,5	3,5	0,5875	82,4125	79,2323913	
24	23	94			14,5875	79,4125	80,4936957	
25	24	73			-7,3375	80,3375	81,755	
26								

Копируем значения «Остатки».

O	P	Q	R	S	T	
	Уровень доверия	0,95				
		Коэффициенты	Среднеквадратическое отклонение	t-статистика	P-значение	Н
	Пересечение	51,48369565217	0,592904893041722	86,832974827	2,1283E-29	50
	X1	1,261304347826	0,041494698684333	30,396758811	1,8302E-19	1,1
	X1	Прогнозное Y	Y	Остатки		
	1	52,745	52,8375	0,0925		
	2	54,00630434783	54,4125	0,4061956522		
	3	55,26760869565	55,4125	0,1448913043		
	4	56,52891304348	57,3375	0,8085869565		
	5	57,7902173913	58,8375	1,0472826087		
	6	59,05152173913	58,4125	-0,639021739		
	7	60,31282608696	60,4125	0,099673913		
	8	61,57413043478	60,3375	-1,236630435		
	9	62,83543478261	62,8375	0,0020652174		
	10	64,09673913043	62,4125	-1,68423913		
	11	65,35804347826	64,4125	-0,945543478		
	12	66,61934782609	69,3375	2,7181521739		
	13	67,88065217391	67,8375	-0,043152174		
	14	69,14195652174	66,4125	-2,729456522		
	15	70,40326086957	70,4125	0,0092391304		
	16	71,66456521739	70,3375	-1,327065217		
	17	72,92586956522	72,8375	-0,088369565		
	18	74,18717391304	76,4125	2,225326087		
	19	75,44847826087	76,4125	0,9640217391		
	20	76,7097826087	76,3375	-0,372282609		
	21	77,97108695652	77,8375	-0,133586957		
	22	79,23239130435	82,4125	3,1801086957		
	23	80,49369565217	79,4125	-1,081195652		
	24	81,755	80,3375	-1,4175		

Вставляем в исходную таблицу. В результате получаем.

	A	B	C	D	E	F	G	H	I
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	
2	1	45			-7,8375	52,8375	52,745	0,0925	
3	2	55			0,5875	54,4125	54,0063043	0,4061956522	
4	3	70	55,75	14,25	14,5875	55,4125	55,2676087	0,1448913043	
5	4	50	57	-7	-7,3375	57,3375	56,528913	0,8085869565	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,7902174	1,0472826087	
7	6	59	59,125	-0,125	0,5875	58,4125	59,0515217	-0,639021739	
8	7	75	60	15	14,5875	60,4125	60,3128261	0,099673913	
9	8	53	61	-8	-7,3375	60,3375	61,5741304	-1,236630435	
10	9	55	62	-7	-7,8375	62,8375	62,8354348	0,0020652174	
11	10	63	63,625	-0,625	0,5875	62,4125	64,0967391	-1,68423913	
12	11	79	65,375	13,625	14,5875	64,4125	65,3580435	-0,945543478	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,6193478	2,7181521739	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,8806522	-0,043152174	
15	14	67	68,625	-1,625	0,5875	66,4125	69,1419565	-2,729456522	
16	15	85	69,375	15,625	14,5875	70,4125	70,4032609	0,0092391304	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,6645652	-1,327065217	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,9258696	-0,088369565	
19	18	77	74,75	2,25	0,5875	76,4125	74,1871739	2,225326087	
20	19	91	76,125	14,875	14,5875	76,4125	75,4484783	0,9640217391	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,7097826	-0,372282609	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971087	-0,133586957	
23	22	83	79,5	3,5	0,5875	82,4125	79,2323913	3,1801086957	
24	23	94			14,5875	79,4125	80,4936957	-1,081195652	
25	24	73			-7,3375	80,3375	81,755	-1,4175	
26									

Сделаем проверку вычислений. Сумма «Сезонной компоненты», «Тренда» и «Случайной компоненты» в каждом квартале должны дать значение «Количества краж».

	A	B	C	D	E	F	G	H	I	J
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	=E2+G2+H2	
3	2	55			0,5875	54,4125	54,0063043	0,4061956522		
4	3	70	55,75	14,25	14,5875	55,4125	55,2676087	0,1448913043		
5	4	50	57	-7	-7,3375	57,3375	56,528913	0,8085869565		
6	5	51	58,125	-7,125	-7,8375	58,8375	57,7902174	1,0472826087		
7	6	59	59,125	-0,125	0,5875	58,4125	59,0515217	-0,639021739		
8	7	75	60	15	14,5875	60,4125	60,3128261	0,099673913		
9	8	53	61	-8	-7,3375	60,3375	61,5741304	-1,236630435		
10	9	55	62	-7	-7,8375	62,8375	62,8354348	0,0020652174		
11	10	63	63,625	-0,625	0,5875	62,4125	64,0967391	-1,68423913		
12	11	79	65,375	13,625	14,5875	64,4125	65,3580435	-0,945543478		
13	12	62	66,5	-4,5	-7,3375	69,3375	66,6193478	2,7181521739		
14	13	60	67,75	-7,75	-7,8375	67,8375	67,8806522	-0,043152174		
15	14	67	68,625	-1,625	0,5875	66,4125	69,1419565	-2,729456522		
16	15	85	69,375	15,625	14,5875	70,4125	70,4032609	0,0092391304		
17	16	63	71,25	-8,25	-7,3375	70,3375	71,6645652	-1,327065217		
18	17	65	73,25	-8,25	-7,8375	72,8375	72,9258696	-0,088369565		
19	18	77	74,75	2,25	0,5875	76,4125	74,1871739	2,225326087		
20	19	91	76,125	14,875	14,5875	76,4125	75,4484783	0,9640217391		
21	20	69	77,5	-8,5	-7,3375	76,3375	76,7097826	-0,372282609		
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971087	-0,133586957		
23	22	83	79,5	3,5	0,5875	82,4125	79,2323913	3,1801086957		
24	23	94			14,5875	79,4125	80,4936957	-1,081195652		
25	24	73			-7,3375	80,3375	81,755	-1,4175		

Получаем те же значения.

	A	B	C	D	E	F	G	H	I	J
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,0063043	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,2676087	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,7902174	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,0515217	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,3128261	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,5741304	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,8354348	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,0967391	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,3580435	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,6193478	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,8806522	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,1419565	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,4032609	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,6645652	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,9258696	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,1871739	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,4484783	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,7097826	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971087	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,2323913	3,1801086957	83	
24	23	94			14,5875	79,4125	80,4936957	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	

Приступим к прогнозу на следующий за отчетным периодом год. Для этого скопируем значения сезонной компоненты, которая повторяется из года в год в 25, 26, 27 и 28 кварталы.

	A	B	C	D	E	F	G	H	I	J
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,0063043	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,2676087	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,7902174	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,0515217	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,3128261	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,5741304	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,8354348	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,0967391	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,3580435	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,6193478	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,8806522	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,1419565	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,4032609	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,6645652	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,9258696	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,1871739	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,4484783	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,7097826	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971087	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,2323913	3,1801086957	83	
24	23	94			14,5875	79,4125	80,4936957	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25		Прогноз на будущий год		-7,8375					
27	26				0,5875					
28	27				14,5875					
29	28				-7,3375					

Рассчитаем тренд в исследуемом году, используя уравнение регрессии. Буквы ячеек со значениями коэффициентов заключим знаком \$ для неизменности при копировании формулы.

	A	B	C	D	E	F	G	H	I	J
13	12	62	66,5	-4,5	-7,3375	69,3375	66,619347826	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,880652174	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,141956522	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,40326087	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,664565217	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,925869565	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,187173913	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,448478261	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,709782609	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971086957	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,232391304	3,1801086957	83	
24	23	94			14,5875	79,4125	80,493695652	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25		Прогноз на будущий год		-7,8375	=B\$40+B\$41*A26				
27	26			0,5875						
28	27			14,5875						
29	28			-7,3375						
30										
31	Дисперсионный анализ (ANOVA)									
32		df	SS	MS	F	Значимость F				
33	Регрессия	1	1829,52195652174	1829,521957	923,9629462	1,83024508633E-19				
34	Остатки	22	43,5617934782608	1,980081522						
35	Итог	23	1873,08375							
36										
37	Уровень доверия	0,95								
38										
39		Коэффициенты	Среднеквадратическое отклонение	t-статистика	P-значение	Нижний 95%	Верхний 95%			
40	Пересечение	51,4836956522	0,592904893041722	86,83297483	2,12835E-29	50,2540861626662	52,713305142			
41	X1	1,26130434783	0,041494698684333	30,39675881	1,83025E-19	1,17524960976386	1,3473590859			
42										

В результате получим.

	A	B	C	D	E	F	G	H	I	
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,006304348	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,267608696	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913043	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,790217391	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,051521739	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,312826087	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,574130435	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,835434783	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,09673913	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,358043478	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,619347826	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,880652174	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,141956522	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,40326087	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,664565217	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,925869565	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,187173913	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,448478261	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,709782609	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971086957	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,232391304	3,1801086957	83	
24	23	94			14,5875	79,4125	80,493695652	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25				-7,8375	83,016304348				
27	26		Прогноз на будущий год		0,5875	84,277608696				
28	27				14,5875	85,538913043				
29	28				-7,3375	86,800217391				

Делаем прогноз по количеству краж в исследуемом году. Данные значения определяем как сумму «Сезонной компоненты» и «Оценки тренда».

	A	B	C	D	E	F	G	H	I	
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,006304348	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,267608696	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913043	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,790217391	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,051521739	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,312826087	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,574130435	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,835434783	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,09673913	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,358043478	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,619347826	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,880652174	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,141956522	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,40326087	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,664565217	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,925869565	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,187173913	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,448478261	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,709782609	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971086957	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,232391304	3,1801086957	83	
24	23	94			14,5875	79,4125	80,493695652	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25	=E26+F26			-7,8375	83,016304348				
27	26		Прогноз на будущий год		0,5875	84,277608696				
28	27				14,5875	85,538913043				
29	28				-7,3375	86,800217391				
30										

И для всего года имеем следующие значения.

	A	B	C	D	E	F	G	H	I	J
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,006304348	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,267608696	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913043	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,790217391	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,051521739	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,312826087	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,574130435	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,835434783	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,09673913	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,358043478	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,619347826	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,880652174	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,141956522	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,40326087	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,664565217	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,925869565	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,187173913	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,448478261	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,709782609	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971086957	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,232391304	3,1801086957	83	
24	23	94			14,5875	79,4125	80,493695652	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25	75,1788043			-7,8375	83,016304348				
27	26	84,8651087	Прогноз на будущий год		0,5875	84,277608696				
28	27	100,126413			14,5875	85,538913043				
29	28	79,4627174			-7,3375	86,800217391				
30										

Округлим полученные значения до целого, выделив ячейки, нажмем на правую клавишу и выберем «Формат ячеек».

Формат ячеек

ЧислаШрифтЭффекты шрифтаВыравниваниеОбрамлениеФонЗащита ячейки

Категория

Все
Особый
Числовой
Процентный
Денежный
Дата
Время
Научный
Дробь
Логический
Текст

Формат

Standard
-1235
-1234,57
-1 235
-1 234,57
-1 234,57
-1 235
-1 235
-1 234,57
-1 234,57
-1234,5678901234600000

Язык

Стандарт - Русский

75

Параметры

Дробная часть: 0

Ведущие нули: 0

☐ Отрицательные числа красным

☐ Разделитель разрядов

Код формата

#

Особый

Справка

Восстановить

Отменить

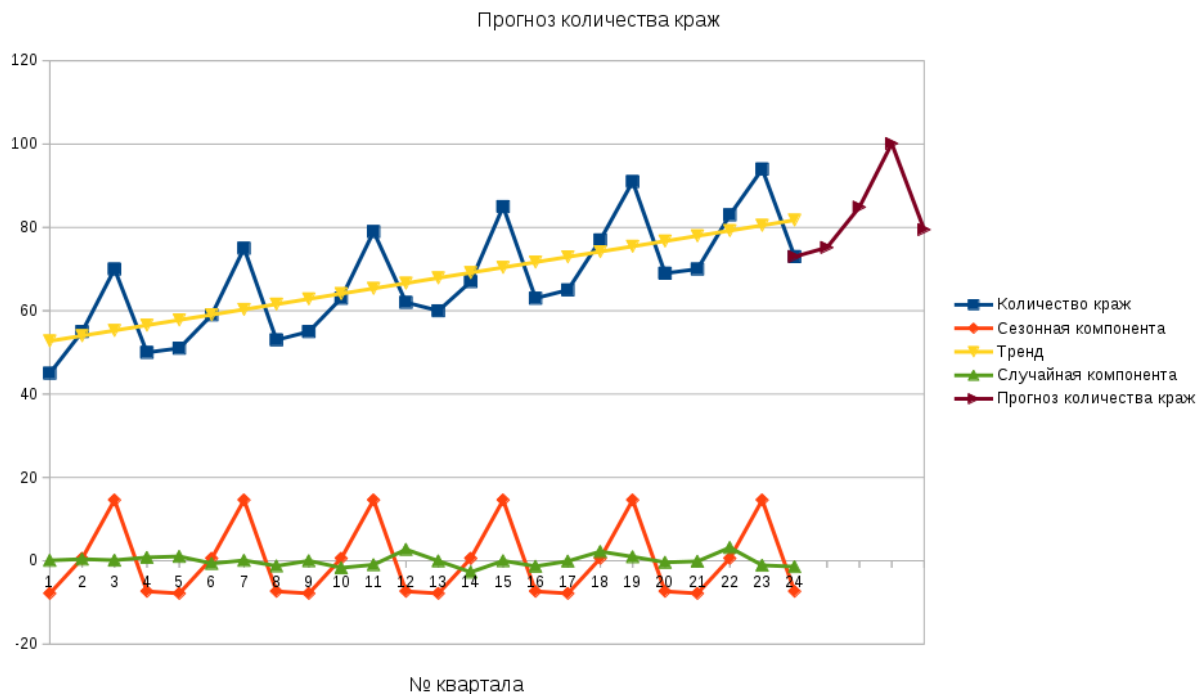
ОК

Окончательно получаем.

	A	B	C	D	E	F	G	H	I	J
1	№ квартала	Количество краж	Скользящая средняя	Оценка сезонной компоненты	Сезонная компонента	Оценка тренда	Тренд	Случайная компонента	Проверка	
2	1	45			-7,8375	52,8375	52,745	0,0925	45	
3	2	55			0,5875	54,4125	54,006304348	0,4061956522	55	
4	3	70	55,75	14,25	14,5875	55,4125	55,267608696	0,1448913043	70	
5	4	50	57	-7	-7,3375	57,3375	56,528913043	0,8085869565	50	
6	5	51	58,125	-7,125	-7,8375	58,8375	57,790217391	1,0472826087	51	
7	6	59	59,125	-0,125	0,5875	58,4125	59,051521739	-0,639021739	59	
8	7	75	60	15	14,5875	60,4125	60,312826087	0,099673913	75	
9	8	53	61	-8	-7,3375	60,3375	61,574130435	-1,236630435	53	
10	9	55	62	-7	-7,8375	62,8375	62,835434783	0,0020652174	55	
11	10	63	63,625	-0,625	0,5875	62,4125	64,09673913	-1,68423913	63	
12	11	79	65,375	13,625	14,5875	64,4125	65,358043478	-0,945543478	79	
13	12	62	66,5	-4,5	-7,3375	69,3375	66,619347826	2,7181521739	62	
14	13	60	67,75	-7,75	-7,8375	67,8375	67,880652174	-0,043152174	60	
15	14	67	68,625	-1,625	0,5875	66,4125	69,141956522	-2,729456522	67	
16	15	85	69,375	15,625	14,5875	70,4125	70,40326087	0,0092391304	85	
17	16	63	71,25	-8,25	-7,3375	70,3375	71,664565217	-1,327065217	63	
18	17	65	73,25	-8,25	-7,8375	72,8375	72,925869565	-0,088369565	65	
19	18	77	74,75	2,25	0,5875	76,4125	74,187173913	2,225326087	77	
20	19	91	76,125	14,875	14,5875	76,4125	75,448478261	0,9640217391	91	
21	20	69	77,5	-8,5	-7,3375	76,3375	76,709782609	-0,372282609	69	
22	21	70	78,625	-8,625	-7,8375	77,8375	77,971086957	-0,133586957	70	
23	22	83	79,5	3,5	0,5875	82,4125	79,232391304	3,1801086957	83	
24	23	94			14,5875	79,4125	80,493695652	-1,081195652	94	
25	24	73			-7,3375	80,3375	81,755	-1,4175	73	
26	25	75	Прогноз на будущий год		-7,8375	83,016304348				
27	26	85			0,5875	84,277608696				
28	27	100			14,5875	85,538913043				
29	28	79			-7,3375	86,800217391				
30										

Согласно сделанного прогноза в первом, втором, третьем и четвертом квартале 7 года будет предположительно совершено 75, 85, 100, 79 краж соответственно.

График распределения количества краж, сезонной компоненты, тренда, случайной компоненты, а так же прогноз количества краж на будущий год имеет вид.



5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Обосновать использование методов математической статистики для анализа данных в правовой области.
2. Каковы основные цели статистического моделирования?
3. Что называют случайной величиной?
4. Что называют генеральной совокупностью?
5. Что называют выборкой наблюдений?
6. Что означает репрезентативность выборки?
7. Какие три типа выборок наблюдений существуют?
8. Как определяется выборочное среднее значение случайной величины?
9. Как определяется выборочная дисперсия случайной величины?
10. Как определяется выборочное среднее квадратическое отклонение случайной величины?
11. Оценкой чего является выборочное среднее квадратическое отклонение?
12. Что такое гистограмма?
13. Как построить гистограмму в программе Calc?
14. С помощью какой функции в программе Calc можно вычислить выборочное среднее значение?
15. С помощью какой функции в программе Calc можно вычислить выборочное среднее квадратическое отклонение?
16. Что такое множественная регрессия?
17. Как записывается уравнение множественной линейной регрессии?
18. Что такое парные коэффициенты корреляции?
19. Как при помощи парных коэффициентов корреляции определить корреляционную связь между факторами?
20. Какие факторы должны быть включены в модель регрессии?
21. Что такое множественный индекс корреляции?
22. Как определяется значимость множественного индекса корреляции?
23. Что называется коэффициентом детерминации?
24. Что такое скорректированный индекс детерминации?
25. Что показывает коэффициент детерминации?
26. Какие недостатки есть у скорректированного индекса детерминации?
27. Что такое аппроксимация?
28. Как определяется средняя ошибка аппроксимации?
29. Что такое уровень значимости?
30. Как определяется значимость коэффициента корреляции?
31. Как оценить значимость уравнения множественной линейной регрессии?
32. В чем заключается содержательный смысл коэффициентов уравнения множественной линейной регрессии?
33. Как вычислить расчетные значения, зная уравнение множественной линейной регрессии?
34. Как сделать прогноз, зная уравнение множественной линейной регрессии?
35. Как сравнить несколько уравнений множественной линейной регрессии и каким образом выбрать одно из них?
36. Как в программе Calc построить график временного ряда?
37. Как определить наличие цикличности в данных?
38. Как вычисляется скользящее среднее для цикличности 4?
39. Как вычисляется скользящее среднее для цикличности 2?

40. Как вычисляется оценка сезонной компоненты?
41. Как определяется оценка тренда модели?
42. Как определяется тренд?
43. Как определяется случайная компонента?
44. Как сделать проверку модели?
45. Как сделать прогноз модели?
46. Как в программе Calc построить график распределения изучаемого явления, сезонной компоненты, тренда, случайной компоненты, а также прогноза?

6. ВАРИАНТЫ ЗАДАНИЙ

Задача 1.

Вариант 1.

Количество и возраст сотрудников, проходящих службу в 404 подразделении Управления «Г» приведены в таблице.

Возраст сотрудника	26	35	39	42	45	50
Количество сотрудников	12	9	11	8	4	1

Необходимо:

1. Построить гистограмму (график) распределения сотрудников по возрасту.
2. Определить средний возраст сотрудников.
3. Сделать вывод о качественном составе данного подразделения (есть ли преемственность поколений и т.п.).

Вариант 2.

После 12 выстрелов по спортивной мишени два стрелка получили результаты, сведенные в таблицу.

Стрелок 1				
Количество выбитых очков	«2»	«5»	«7»	«10»
Количество попаданий	2	5	4	1
Стрелок 2				
Количество выбитых очков	«4»	«5»	«6»	«7»
Количество попаданий	1	4	4	1

Необходимо:

1. Определить среднее значений выбитых очков и разброс их около среднего значения каждого стрелка.
2. Сделать вывод о качестве подготовки стрелков.

Вариант 3.

Количество и возраст сотрудников, проходящих службу в 404 подразделении Управления «Д» приведены в таблице.

Возраст сотрудника	33	37	40	41	42	43
Количество сотрудников	1	8	1	23	3	2

Необходимо:

1. Построить гистограмму (график) распределения сотрудников по возрасту.
2. Определить средний возраст сотрудников.
3. Сделать вывод о качественном составе данного подразделения (есть ли преобладание поколений и т.п.).

Вариант 4.

После 12 выстрелов по спортивной мишени два стрелка получили результаты, сведенные в таблицу.

Стрелок 1				
Количество выбитых очков	«2»	«5»	«7»	«10»
Количество попаданий	1	5	4	2
Стрелок 2				
Количество выбитых очков	«4»	«5»	«6»	«7»
Количество попаданий	2	3	4	1

Необходимо:

1. Определить среднее значений выбитых очков и разброс их около среднего значения каждого стрелка.
2. Сделать вывод о качестве подготовки стрелков.

Вариант 5.

При изучении дисциплины «Правовая статистика» студент Иванов получил следующие оценки: 3, 2, 3, 5, 3, 5, 3, 2, 3, 4, 3, 3, 3, 5, 5, 4, 2, 4, 2, 4, 5, 5, 2, 3, 4, 2, 5, 4, 5, 4, 2, 4, 3, 3, 2, 3, 2, 4, 4, 3, 2.

Необходимо:

1. Построить гистограмму (график) распределения оценок.
2. Определить средний бал студента, найти разброс оценок около среднего бала.
3. Сделать вывод о качестве освоения дисциплины.

Вариант 6.

Количество и возраст сотрудников, проходящих службу в 404 подразделении Управления «Е» приведены в таблице.

Возраст сотрудника	22	27	39	42	41	42
Количество сотрудников	25	2	1	8	3	1

Необходимо:

1. Построить гистограмму (график) распределения сотрудников по возрасту.
2. Определить средний возраст сотрудников.
3. Сделать вывод о качественном составе данного подразделения (есть ли преобладание поколений и т.п.).

Вариант 7.

После 12 выстрелов по спортивной мишени два стрелка получили результаты, сведенные в таблицу.

Стрелок 1				
Количество выбитых очков	«2»	«5»	«7»	«10»
Количество попаданий	1	4	4	1
Стрелок 2				
Количество выбитых очков	«4»	«5»	«6»	«7»
Количество попаданий	2	5	4	1

Необходимо:

1. Определить среднее значений выбитых очков и разброс их около среднего значения каждого стрелка.
2. Сделать вывод о качестве подготовки стрелков.

Вариант 8.

Количество и возраст сотрудников, проходящих службу в 404 подразделении Управления «Ж» приведены в таблице.

Возраст сотрудника	21	22	23	24	25	43
Количество сотрудников	12	5	11	5	13	2

Необходимо:

1. Построить гистограмму (график) распределения сотрудников по возрасту.

2. Определить средний возраст сотрудников.
3. Сделать вывод о качественном составе данного подразделения (есть ли преобладание поколений и т.п.).

Вариант 9.

После 12 выстрелов по спортивной мишени два стрелка получили результаты, сведенные в таблицу.

Стрелок 1				
Количество выбитых очков	«2»	«5»	«7»	«10»
Количество попаданий	1	5	4	2
Стрелок 2				
Количество выбитых очков	«3»	«6»	«7»	«8»
Количество попаданий	2	3	4	1

Необходимо:

1. Определить среднее значений выбитых очков и разброс их около среднего значения каждого стрелка.
2. Сделать вывод о качестве подготовки стрелков.

Вариант 0.

При изучении дисциплины «Правовая статистика» студент Иванов получил следующие оценки: 3, 3, 3, 5, 3, 5, 3, 3, 3, 4, 3, 2, 3, 5, 5, 4, 3, 4, 3, 4, 5, 5, 3, 3, 4, 3, 5, 4, 5, 4, 3, 4, 3, 3, 3, 3, 3, 4, 4, 3, 3.

Необходимо:

1. Построить гистограмму (график) распределения оценок.
2. Определить средний бал студента, найти разброс оценок около среднего бала.
3. Сделать вывод о качестве освоения дисциплины.

Задача 2. В городе М-ске были получены статистические данные по 20 опорным пунктам участковых инспекторов полиции за отчетный период, которые сведены в таблицу.

Необходимо:

1. Определить существование зависимости количества преступлений от численности населения, количества осужденных и количества раскрытых преступлений по городу в целом.

2. Сделать прогноз оценки количества преступлений на 11 участке, если количество осужденных уменьшится на 10, а количество раскрытых преступлений увеличиться на 15, а численность населения увеличиться на 20000 человек.

Вариант 1.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	55	65000	25	46
2.	67	62000	38	59
3.	78	73000	43	69
4.	86	84000	47	79
5.	67	72000	38	62
6.	90	81000	49	83
7.	45	55000	27	37
8.	67	73000	38	58
9.	66	70000	37	62
10.	43	45000	26	34
11.	23	21000	16	16
12.	23	26000	16	14
13.	56	64000	32	47
14.	90	89000	49	83
15.	56	66000	32	47
16.	67	68000	38	65
17.	68	78000	38	59
18.	32	40000	20	29
19.	12	17000	10	7
20.	34	25000	21	33

Вариант 2.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	22	65000	10	12
2.	88	62000	49	80
3.	32	73000	19	29
4.	78	84000	40	68
5.	2	72000	3	1
6.	10	81000	4	5

7.	56	55000	23	50
8.	55	73000	30	45
9.	64	70000	29	54
10.	30	45000	16	29
11.	12	21000	11	9
12.	15	26000	10	8
13.	11	64000	6	7
14.	69	89000	40	61
15.	57	66000	29	54
16.	43	68000	24	40
17.	70	78000	36	69
18.	39	40000	24	31
19.	61	17000	32	58
20.	22	25000	10	12

Вариант 3.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	60	65000	35	59
2.	34	62000	22	24
3.	56	73000	32	54
4.	30	84000	16	20
5.	56	72000	30	51
6.	34	81000	20	33
7.	80	55000	41	75
8.	48	73000	27	38
9.	28	70000	18	26
10.	58	45000	31	54
11.	68	21000	37	65
12.	38	26000	23	34
13.	32	64000	17	30
14.	18	89000	12	16
15.	52	66000	31	48
16.	78	68000	40	70
17.	98	78000	53	92
18.	96	40000	51	90
19.	92	17000	51	85
20.	94	25000	52	90

Вариант 4.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	54	65000	28	51
2.	65	62000	35	60

3.	85	73000	45	81
4.	68	84000	36	62
5.	95	72000	50	92
6.	53	81000	29	46
7.	67	55000	36	62
8.	61	73000	33	59
9.	59	70000	32	53
10.	74	45000	39	73
11.	58	21000	31	50
12.	43	26000	24	40
13.	76	64000	40	71
14.	80	89000	42	76
15.	72	66000	38	70
16.	93	68000	49	85
17.	56	78000	30	55
18.	74	40000	39	71
19.	34	17000	19	29
20.	54	25000	28	51

Вариант 5.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	19	65000	8	9
2.	31	62000	10	25
3.	82	73000	40	77
4.	95	84000	41	90
5.	27	72000	19	20
6.	91	81000	40	83
7.	39	55000	14	35
8.	58	73000	24	57
9.	47	70000	19	43
10.	81	45000	35	75
11.	27	21000	12	21
12.	35	26000	14	34
13.	80	64000	35	80
14.	69	89000	30	64
15.	24	66000	10	16
16.	41	68000	17	37
17.	49	78000	18	40
18.	63	40000	28	59
19.	29	17000	12	21
20.	71	25000	34	64

Вариант 6.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	12	7000	4	6
2.	46	41000	21	41
3.	89	84000	43	80
4.	45	40000	21	43
5.	76	71000	32	74
6.	89	84000	43	83
7.	67	62000	31	64
8.	34	29000	15	30
9.	67	62000	31	66
10.	87	82000	41	83
11.	98	93000	47	95
12.	32	27000	14	32
13.	45	4000	21	40
14.	78	73000	37	70
15.	89	84000	43	86
16.	65	60000	30	60
17.	45	40000	21	44
18.	78	73000	37	76
19.	76	71000	31	70
20.	34	29000	15	33

Вариант 7.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	43	48000	25	37
2.	23	26000	38	15
3.	54	61000	43	44
4.	67	71000	47	60
5.	97	99000	38	90
6.	12	17000	49	3
7.	46	52000	27	40
8.	65	70000	38	55
9.	17	22000	37	7
10.	19	24000	26	14
11.	92	97000	16	84
12.	31	39000	16	23
13.	76	81000	32	70
14.	14	17000	49	5
15.	34	39000	32	24
16.	78	82000	38	70
17.	41	46000	38	31

18.	38	40000	20	30
19.	61	68000	10	52
20.	87	89000	21	80

Вариант 8.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	10	7000	25	8
2.	15	17000	38	12
3.	25	28000	43	15
4.	23	22000	47	21
5.	11	2000	38	10
6.	14	16000	49	11
7.	22	24000	27	19
8.	30	20000	38	20
9.	24	21000	37	19
10.	21	25000	26	18
11.	18	23000	16	14
12.	13	21000	16	10
13.	19	23000	32	10
14.	16	20000	49	9
15.	33	37000	32	25
16.	29	33000	38	20
17.	26	30000	38	21
18.	21	25000	20	20
19.	17	21000	10	10
20.	20	24000	21	15

Вариант 9.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	56	58000	25	66
2.	23	20000	38	25
3.	13	15000	43	17
4.	54	51000	47	56
5.	21	16000	38	26
6.	89	94000	49	93
7.	27	31000	27	30
8.	36	32000	38	39
9.	91	88000	37	94
10.	55	64000	26	57
11.	72	69000	16	77
12.	33	38000	16	35
13.	50	40000	32	58
14.	13	15000	49	15

15.	75	82000	32	78
16.	11	13000	38	16
17.	18	25000	38	25
18.	49	54000	20	52
19.	61	62000	10	62
20.	44	53000	21	46

Вариант 0.

№ п/п	Количество преступлений	Численность населения	Количество осужденных	Количество раскрытых преступлений
1.	23	65000	12	23
2.	34	62000	19	33
3.	54	73000	30	52
4.	21	84000	14	18
5.	66	72000	38	62
6.	74	81000	36	69
7.	79	55000	42	73
8.	45	73000	26	38
9.	36	70000	22	27
10.	37	45000	24	27
11.	48	21000	23	47
12.	27	26000	13	25
13.	22	64000	14	21
14.	24	89000	15	22
15.	93	66000	46	83
16.	92	68000	49	90
17.	91	78000	44	89
18.	46	40000	22	44
19.	47	17000	27	43
20.	31	25000	21	29

Задача 3. За 7 лет работы подразделения полиции были собраны ежеквартальные данные о количестве краж, совершенных на подконтрольной территории, значения которых сведены в таблицу.

Необходимо: на основании имеющихся данных сделать прогноз оценки количества краж на год, следующий за отчетным периодом.

Вариант 1.

Год	№ квартала	Количество краж
1	1	60
	2	70
	3	85
	4	65
2	5	66
	6	74
	7	90
	8	68
3	9	70
	10	78
	11	94
	12	77
4	13	75
	14	82
	15	100
	16	78
5	17	80
	18	92
	19	106
	20	84
6	21	85
	22	98
	23	109
	24	88
7	25	90
	26	100
	27	115
	28	94

Вариант 2.

Год	№ квартала	Количество краж
1	1	58
	2	68
	3	83
	4	63
2	5	64
	6	72
	7	88
	8	66
3	9	68
	10	76
	11	92
	12	75
4	13	73
	14	80
	15	98
	16	76
5	17	78
	18	90
	19	104
	20	82
6	21	83
	22	96
	23	107
	24	86
7	25	88
	26	98
	27	113
	28	92

Вариант 3.

Год	№ квартала	Количество краж
1	1	67
	2	77
	3	92
	4	72

Вариант 4.

Год	№ квартала	Количество краж
1	1	82
	2	92
	3	107
	4	87

2	5	73
	6	81
	7	97
	8	75
3	9	77
	10	85
	11	101
	12	84
4	13	82
	14	89
	15	107
	16	85
5	17	87
	18	99
	19	113
	20	91
6	21	92
	22	105
	23	116
	24	95
7	25	97
	26	107
	27	122
	28	101

2	5	88
	6	96
	7	112
	8	90
3	9	92
	10	100
	11	116
	12	99
4	13	97
	14	104
	15	122
	16	100
5	17	102
	18	114
	19	128
	20	106
6	21	107
	22	120
	23	131
	24	110
7	25	112
	26	122
	27	137
	28	116

Вариант 5.

Год	№ квартала	Количество краж
1	1	41
	2	51
	3	66
	4	46
2	5	47
	6	55
	7	71
	8	49
3	9	51
	10	59
	11	75
	12	58
4	13	56
	14	63
	15	81
	16	59
5	17	61
	18	73
	19	87
	20	65

Вариант 6.

Год	№ квартала	Количество краж
1	1	138
	2	148
	3	163
	4	143
2	5	144
	6	152
	7	168
	8	146
3	9	148
	10	156
	11	172
	12	155
4	13	153
	14	160
	15	178
	16	156
5	17	158
	18	170
	19	184
	20	162

6	21	66
	22	79
	23	90
	24	69
7	25	71
	26	81
	27	96
	28	75

Вариант 7.

Год	№ квартала	Количество краж
1	1	34
	2	44
	3	59
	4	39
2	5	40
	6	48
	7	64
	8	42
3	9	44
	10	52
	11	68
	12	51
4	13	49
	14	56
	15	74
	16	52
5	17	54
	18	66
	19	80
	20	58
6	21	59
	22	72
	23	83
	24	62
7	25	64
	26	74
	27	89
	28	68

Вариант 9.

Год	№ квартала	Количество краж
1	1	102
	2	112
	3	127

6	21	163
	22	176
	23	187
	24	166
7	25	168
	26	178
	27	193
	28	172

Вариант 8.

Год	№ квартала	Количество краж
1	1	152
	2	162
	3	177
	4	157
2	5	158
	6	166
	7	182
	8	160
3	9	162
	10	170
	11	186
	12	169
4	13	167
	14	174
	15	192
	16	170
5	17	172
	18	184
	19	198
	20	176
6	21	177
	22	190
	23	201
	24	180
7	25	182
	26	192
	27	207
	28	186

Вариант 0.

Год	№ квартала	Количество краж
1	1	33
	2	43
	3	58

	4	107
2	5	108
	6	116
	7	132
	8	110
3	9	112
	10	120
	11	136
	12	119
4	13	117
	14	124
	15	142
	16	120
5	17	122
	18	134
	19	148
	20	126
6	21	127
	22	140
	23	151
	24	130
7	25	132
	26	142
	27	157
	28	136

	4	38
2	5	39
	6	47
	7	63
	8	41
3	9	43
	10	51
	11	67
	12	50
4	13	48
	14	55
	15	73
	16	51
5	17	53
	18	65
	19	79
	20	57
6	21	58
	22	71
	23	82
	24	61
7	25	63
	26	73
	27	88
	28	67

ЛИТЕРАТУРА

1. Правовая статистика: учебник для студентов вузов, обучающихся по специальности «Юриспруденция», для курсантов и слушателей образовательных учреждений МВД / Под ред. В.С. Лялина, А.В. Симоненко. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2013. – 255 с.
2. Правовая статистика: учебник для студентов вузов, обучающихся по специальности 030501 «Юриспруденция», для курсантов и слушателей образовательных учреждений МВД / Под ред. С.Я. Казанцева, С.Я. Лебедева. –М.: ЮНИТИ-ДАНА, Закон и право, 2008. – 255 с.
3. Лялин В.С. Правовая статистика: учебник / В.С. Лялин, Е.А. Костыря – СПб.: ИВЭСЭП, Знание, 2006. – 235 с.
4. Андриюшечкина И.Н. Судебная статистика: учебно-практическое пособие / И.Н. Андриюшечкина, Е.А. Ковалёв. – М.: Проспект, 2014. – 392 с.
5. Просветов Г.И. Математика для гуманитариев: задачи и решения: учебно-практическое пособие. – М.: Издательство «Альфа-Пресс», 2008. – 320 с.
6. Эконометрика / под ред. И.И. Елисеевой. – М.: Проспект, 2011. - 288 с.
7. Кремер Н.Ш. Эконометрика / Н.Ш. Кремер, Б.А. Путко / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2007. – 311 с.
8. Кремер Н.Ш. Теория вероятностей и математическая статистика. – М.: ЮНИТИ-ДАНА, 2009. – 551 с.

Учебное издание

Ольга Юрьевна Данилова,
кандидат физико-математических наук, доцент;
Владимир Александрович Родин,
доктор физико-математических наук, профессор;
Сергей Владимирович Синегубов,
кандидат технических наук, доцент

ПРАВОВАЯ СТАТИСТИКА

Учебное пособие

В авторской редакции.
Компьютерная верстка С.А. Телковой.
Объем 8 Мб

Воронежский институт МВД России
394065 Воронеж, просп. Патриотов, 53