ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ

«Университет «Дубна»

ИНСТИТУТ СИСТЕМНОГО АНАЛИЗА И УПРАВЛЕНИЯ

Кафедра системного анализа и управления

КУРСОВАЯ РАБОТА

по дисциплине

«Теория вероятностей и математическая статистика»

Исследование корреляционной зависимости по выборке

ТЕМА: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2013

Выполнил: студент группы \_\_\_\_\_\_\_\_\_\_\_\_

Ильин Михаил Алексеевич

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(Ф.И.О.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись студента)

**Руководители**:

доцент Копылова Татьяна Валерьевна

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись руководителя)

Дубна, 2021

**ОГЛАВЛЕНИЕ**

[1. Введение 3](#_Toc89912643)

[2. Постановка задачи 4](#_Toc89912644)

[3. Теоритическая часть 5](#_Toc89912645)

[1. Основные теоретические понятия 5](#_Toc89912646)

[2. Метод наименьших квадратов 7](#_Toc89912647)

[3. Проверка гипотезы о значимости выборочного коэффициента корреляции 9](#_Toc89912648)

[4. Практическая часть 10](#_Toc89912649)

[5. Выводы 14](#_Toc89912650)

[6. Список литературы 15](#_Toc89912651)

# Введение

Математическая статистика — это раздел математики, посвященный методам анализа данных, преимущественно вероятностной природы. Она занимается систематизацией, обработкой и использованием статистических данных для теоретических и практических выводов. Во многих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала, выборки генеральной совокупности.

Цель статистического анализа — исследование свойств случайной величины. Для этого приходится несколько раз измерять значения изучаемой случайной величины. Полученная группа значений рассматривается как выборка из гипотетической генеральной совокупности [3].

Во время статистических наблюдений для каждого объекта в ряде случаев можно измерить значение нескольких признаков. Таким образом, получается многомерная выборка. Если многомерную выборку обработать по значениям отдельного признака, то получится обычная обработка одномерной выборки.

Смысл обработки многомерных выборок состоит в том, чтобы установить связь между признаками. Связи между ними могут быть функциональными, то есть каждому значению одной величины соответствует определенное значение другой величины.

Связь между случайными величинами часто носит случайный характер. Она называется статистической, если изменение одной величины вызывает изменение распределения другой величины. Если среднее значение одной случайной величины функционально зависит от значения другой случайной величины, то такая статистическая зависимость называется корреляционной.

# Постановка задачи

Представлены данные выборки:

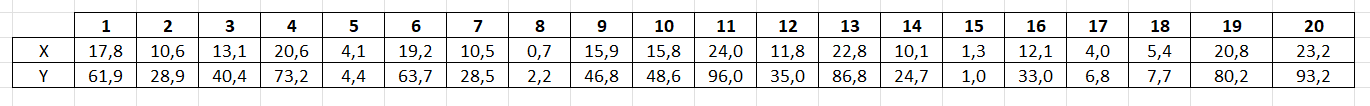


Рис. 1. Исходные данные. Вариант №5.

1. Изучить соответствующий теоретический материал по конспекту лекций и учебнику.
2. Провести статистический анализ данных для переменных *X* и *Y*. Найти выборочные средние, дисперсии и среднеквадратические отклонения для *X* и *Y* по отдельности.
3. Найти ковариацию *Сov (X,Y)*.
4. Найти коэффициент корреляции *X* и *Y*.
5. Найти по выборке уравнение линейной регрессии (*Y* как функцию *X*) по методу наименьших квадратов.
6. Построить графики, изображающие данные выборки и найденные функции регрессии.
7. Проверить гипотезу о значимости выборочного коэффициента корреляции при заданном уровне значимости *α* = 0.01.
8. Проанализировать полученные результаты и сделать выводы.

# Теоритическая часть

## Основные теоретические понятия

Объемом совокупности *n* называют число объектов этой совокупности.

Выборкой называется совокупность случайно отобранных объектов из генеральной совокупности.

Варианты ⸻ это отслеживаемые значения. Вариационный ряд ⸻ последовательность вариантов в возрастающем порядке.

Среднее арифметическое (выборочное среднее) — число, равное сумме всех чисел множества, делённой на их количество. Если все значения *x1, x2, .…, xn* признака выборки объема *n* различны, то:

Дисперсией вариационного ряда называется среднее арифметическое квадратов отклонения вариантов от их средней арифметической:

Выборочное средне квадратическое отклонение  — среднее арифметические абсолютной величины отклонение вариантов от их средней арифметической:

Ковариация (корреляционный момент, ковариационный момент) — мера линейной зависимости двух случайных величин:

Коэффициент корреляции — это статистическая мера, которая вычисляет силу связи между относительными движениями двух переменных. Значения коэффициента корреляции находятся в диапазоне от -1.0 до 1.0. Чем ближе к единице абсолютное значение коэффициента, тем сильнее линейная связь между переменными:

Регрессия — зависимость среднего значения какой-либо величины от некоторой другой величины или нескольких величин.

По форме зависимости различают:

* Линейную регрессию, которая выражается уравнением прямой:

*Y = aX + b,*

* Нелинейную (параболическую):

*Y = PX2 + QX + R.*

## Метод наименьших квадратов

Для исследования линейной регрессии определим коэффициенты линейной функции методом наименьших квадратов. Для этого составим сумму:

Главное — это найти такие коэффициенты линейной зависимости, при которых значение (7) будет наименьшим. Иначе говоря, при определенных значениях *a* и *b* сумма квадратов отклонений представленных данных от получившейся прямой будет иметь минимальное значение. В этом и состоит смысл метода наименьших квадратов.

Для того чтобы эта сумма была минимальной, необходимо, чтобы ее частные производные по параметрам *A* и *B* были равны нулю:

Для решения системы уравнений можно использовать любые методы, например, подстановку или метод Крамера. В результате у нас должны получиться формулы, с помощью которых вычисляются коэффициенты по методу наименьших квадратов:

Выразим *a* и *b*:

## Проверка гипотезы о значимости выборочного коэффициента корреляции

Поскольку выборка отобрана случайно, то нельзя заключить, что коэффициент корреляции генеральной совокупности *r* также отличен от нуля. Возникает необходимость при данном уровне значимости *α* проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе .

В качестве критерия проверки нулевой гипотезы применяют случайную величину

Величина *T* при справедливости нулевой гипотезы имеет распределение Стьюдента с степенями свободы. Поэтому вычисляется эмпирическое значение критерия и по таблице критических точек распределения Стьюдента по выбранному уровню значимости α и числу степеней свободы находят критическую точку

Если, то нулевую гипотезу отвергают, и выборочный коэффициент корреляции значимо отличается от нуля, а *X* и *Y* коррелированы, т.е. связаны линейной зависимостью.

Если то нет оснований отвергать нулевую гипотезу и говорят, что выборочный коэффициент корреляции незначим, а *X* и *Y* некоррелированные, т.е. не связаны линейной зависимостью.

# Практическая часть

Для понимания связи между выбранными параметрами необходимо построить диаграмму рассеивания и понять ситуацию в целом (рис. 2.1).

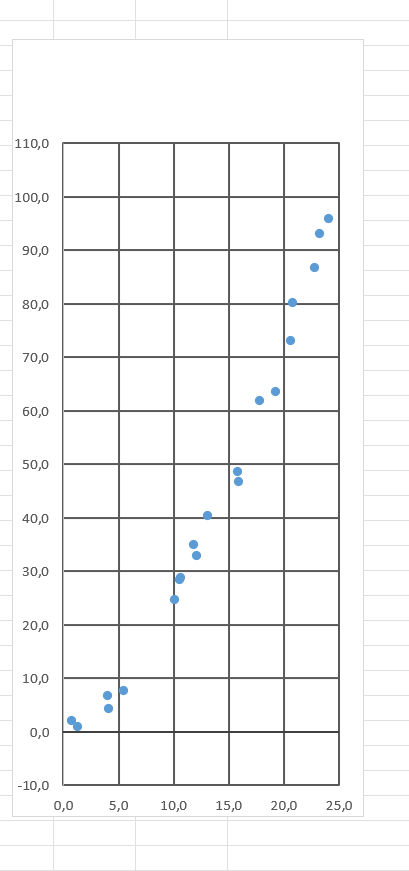


Рис. 2.1. Диаграмма рассеивания.

А также сделаем вариационным исходный ряд, который будет использоваться далее (рис. 2.2).



Рис. 2.2. Вариационный ряд

С помощью исходных данных определим значения выборочного среднего, дисперсии, среднеквадратичного отклонения, ковариации и коэффициента корреляции для *X* и *Y* (рис. 3).

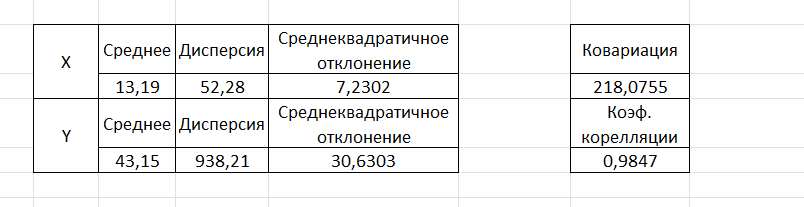


Рис. 3. Результаты вычислений.

При помощи метода наименьших квадратов найдем коэффициенты *a* и *b*, предварительно проведя замену в (8):

Таким образом получаем значения *a* = 4,172 и *b* = –11,874.

Уравнение линии регрессии имеет вид *y* = 4,172*x* – 11,874.

Отобразим линию регрессии на диаграмме рассеивания (рис. 4).

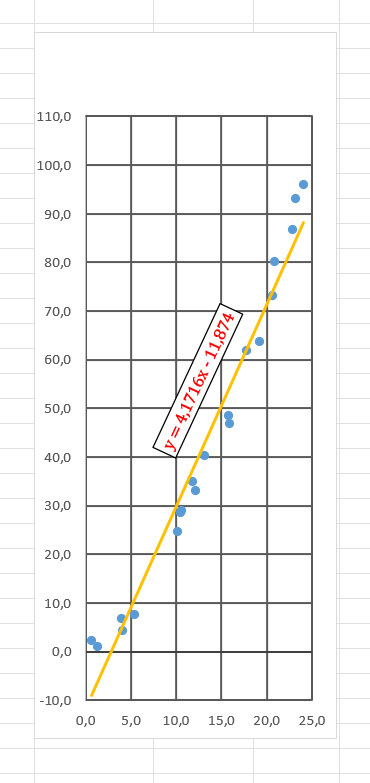


Рис. 4. Диаграмма рассеивания с отображённой линией регрессии.

Проверим нулевую гипотезу о равенстве нулю генерального коэффициента корреляции необходимость при данном уровне значимости *α* = 0,01 при конкурирующей гипотезе .

,

.

Поскольку , гипотеза *H0* отвергается. Выборочный коэффициент корреляции значительно отличается от 0, следовательно, *X* и *Y* коррелированы.

# 5. Выводы

В результате завершения процесса работы над курсовой был проведён статистический анализ предоставленной выборки пар чисел *X* и *Y.*

В данной курсовой работе были вычислены основные числовые характеристики выборок по *X* и *Y*. В процессе выполнения работы был изучен метод наименьших квадратов, с помощью которого были найдены коэффициенты линейной регрессии, построена диаграмма рассеивания с ней. Так же были найдены ковариация и коэффициент корреляции, с помощью которых мы поняли, что они действительно важны при статистическом анализе данных, что они помогают понять поведение случайных величин и как они связаны между собой.

В результате проведенной работы были закреплены теоретические знания и приобретены практические навыки работы со статистиками, также мы узнали, что такое гипотеза о значимости выборочного коэффициента корреляции, которая даёт нам предполагать существование зависимости между генеральной и выборочной совокупностью.

# 6. Список литературы

1. Кремер Н.Ш. Теория вероятностей и математическая статистика — М.: ЮНИТИ-ДАНА, 2006. — глава 12 "Корреляционный анализ".
2. Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1998. 3. — глава 14, параграфы 18-21 (корреляция в теории вероятностей); глава 18, параграфы 1, 3, 4, 7, 14 (элементы теории корреляции в математической статистике); глава 19, параграф 22 (проверка гипотезы о значимости выборочного коэффициента корреляции).
3. Сазонов В.Ф. Математическая статистика, 2009. URL: https://kineziolog.su/content/1-matematicheskaya-statistika-vvedenie (дата обращения: 07.12.2021).