**ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ**

**ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ**

**«УНИВЕРСИТЕТ «ДУБНА»**

(государственный университет «Дубна»)

Кафедра высшей математики

**КУРСОВАЯ РАБОТА**

**по дисциплине (модулю)**

**«ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»**

тема: «Исследование корреляционной зависимости по выборке»

**Выполнил**: студент гр. {group} \_

\_ {name} \_

\_ \_

**Руководитель:**

\_ {teacher} \_

Дата защиты:\_ \_

Оценка: \_ \_

\_ \_

**Дубна {year} г.**

# Введение

Данная работа посвящена исследованию корреляционной зависимости по выборке, построению графика уравнений регрессии и проверке гипотезы о значимости найденного коэффициента корреляции. В ходе работы будет проведен статистический анализ данных для переменных *X* и *Y*.

# Задание на работу

Дана выборка, состоящая из пар чисел (), .

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

# Оглавление

[Введение 2](#_Toc466317383)

[Задание на работу 2](#_Toc466317384)

[Оглавление 3](#_Toc466317385)

[Теоретическая часть 4](#_Toc466317386)

[Практическая часть 7](#_Toc466317387)

[Заключение 9](#_Toc466317388)

[Список литературы 9](#_Toc466317389)

# Теоретическая часть

Приведем основные понятия из курса теории вероятностей и математической статистики, которые будут использоваться в данной работе.

*Выборочной совокупностью* или просто *выборкой* называют совокупность случайно отобранных объектов.

*Математическая статистика* – наука, которая занимается обработкой данных. Во многих своих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надежность и точность выводов, делаемых на основании ограниченного статистического материала (выборки).

*Генеральной совокупностью* называют совокупность объектов, из которых производится выборка.

*Выборочной совокупностью (выборкой)* называют совокупность случайно отобранных объектов.

*Математическим ожиданием* *M(X)* или *выборочным средним* дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности и рассчитывается по формуле:

*Дисперсией* ** дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания и рассчитывается по формуле:

*Средним квадратическим* *отклонением* случайной величины *X* называют квадратный корень из дисперсии:

Ковариацией еще корреляционным моментом *Cov(X, Y)* случайных величин *X* и *Y* называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий:

Ковариацию называют еще вторым смешанным центральным моментом случайных величин *X* и *Y*. Для ковариации *X* и *Y* используются также обозначения , .

*Корреляция* в математической статистике – вероятностная (статистическая) зависимость между величинами, не имеющая, вообще говоря, строго функционального характера.

*Коэффициентом корреляции* ** случайных величин *X* и *Y* называют отношение корреляционного момента к произведению средних квадратичных отклонений этих величин:

Методом наименьших квадратов называется метод нахождения приближения, основанный на минимизации суммы квадратов отклонений приближенных значений от выборочных.

*, где ,*

*,*

*,*

*,*

*,*

*,*

*.*

Пусть двумерная генеральная совокупность *(X, Y)* распределена нормально. Из этой совокупности извлечена выборка объема *n* и по ней найден выборочный коэффициент корреляции *,* который оказался отличным от нуля. Так как выборка отобрана случайно, то еще нельзя заключить, что коэффициент корреляции генеральной совокупности также отличен от нуля. В конечном счете нас интересует именно этот коэффициент, поэтому возникает необходимость при заданном уровне значимости α *проверить нулевую гипотезу* о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе *.*

Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе , надо вычислить наблюдаемое значение критерия:

*;*

и по таблице Стьюдента, по заданному уровню значимости и числу степеней свободынайти критическую точку для двусторонней критической области.

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

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

# Практическая часть

Дана выборка зависимости числа *Y* от числа *X*. Объем выборки .

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | {x1} | {x2} | {x3} | {x4} | {x5} | {x6} | {x7} | {x8} | {x9} | {x10} |
| *Y* | {y1} | {y2} | {y3} | {y4} | {y5} | {y6} | {y7} | {y8} | {y9} | {y10} |
| *X* | {x11} | {x12} | {x13} | {x14} | {x15} | {x16} | {x17} | {x18} | {x19} | {x20} |
| *Y* | {y11} | {y12} | {y13} | {y14} | {y15} | {y16} | {y17} | {y18} | {y19} | {y20} |

Таблица №1. Выборка.

Основные характеристики выборки.

**Измерения по X:**

Математическое ожидание — ;

Выборочная дисперсия —;

Среднеквадратическое отклонение — .

**Измерения по *Y*:**

Математическое ожидание —;

Дисперсия —;

Среднеквадратическое отклонение —.

**Измерения коэффициента ковариации и коэффициента корреляции:**

**ДВА ВАРИАНТА корреляция=0 и |корреляция| стремится к 1**

Полученное значение коэффициента корреляции характеризует обратную связь между исследуемыми величинами, так как значение отрицательно. Так как принимает значение близкое к , то *X* и *Y* связаны линейной функциональной зависимостью.

В ходе измерений были получены коэффициенты для графика, изображающего данные выборки и найденную функцию регрессии.

,

,

*.*

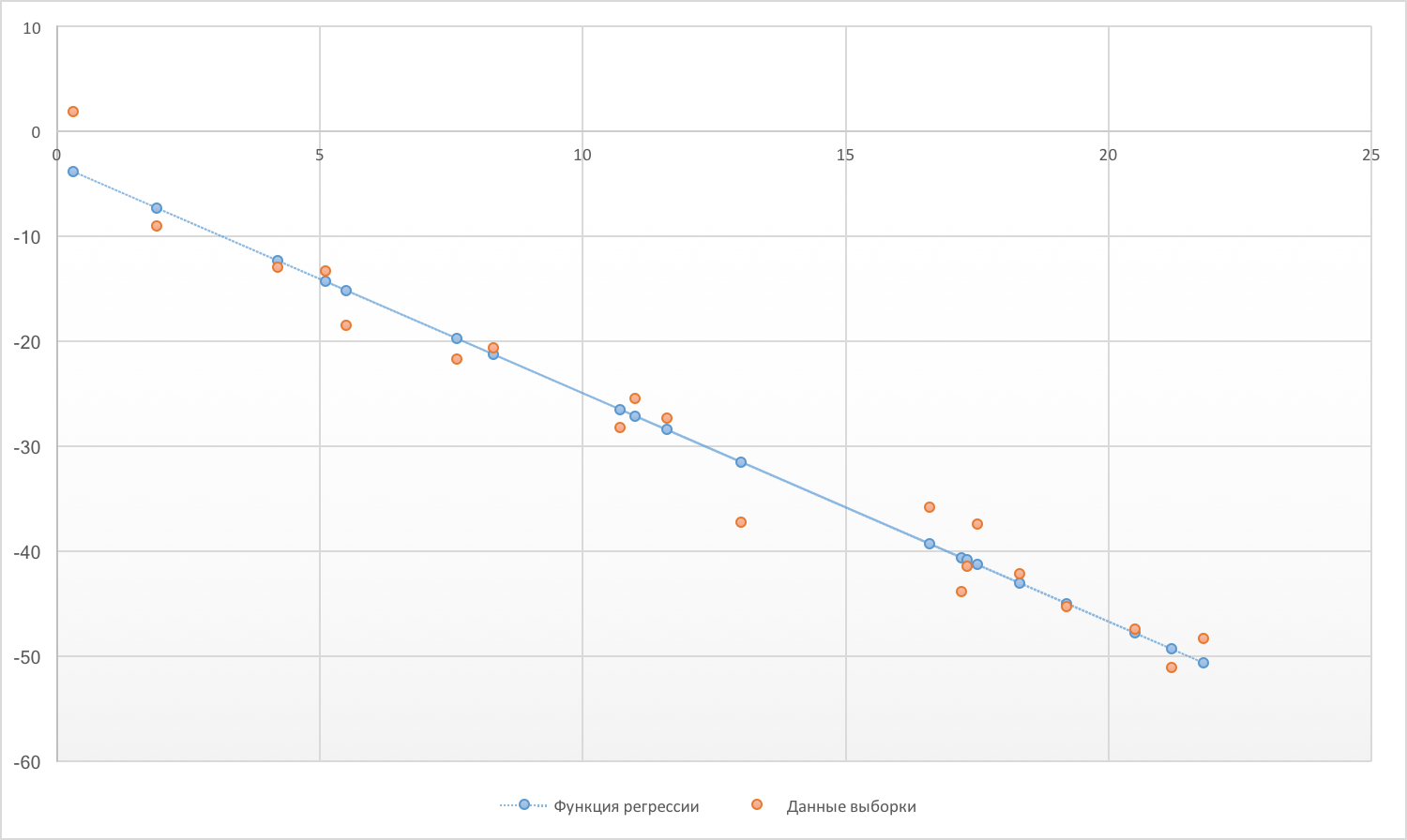


Рисунок №1. Диаграмма с данными выборки и найденной функцией регрессии.

Поскольку оценка коэффициента корреляции вычислена на конечной выборке, и поэтому может отклоняться от своего генерального значения, необходимо проверить значимость коэффициента корреляции. Проверка производится с помощью -критерия:

*= {tDelta}.*

Случайная величина  следует t-распределению Стьюдента и по таблице t-распределения необходимо найти критическое значение критерия при заданном уровне значимости α. Если вычисленное  по модулю окажется меньше чем , то зависимости между случайными величинами X и Y нет. В противном случае, экспериментальные данные не противоречат гипотезе о зависимости случайных величин.

.

**В зависимости от pValue решаем отвергаем ли мы гипотезу**

Абсолютное значение -критерия не меньше критического , , следовательно **экспериментальные данные, с вероятностью** , **не противоречат гипотезе** о зависимости случайных величин X и Y.

Так как то мы отвергаем нулевую гипотезу и говорим, что X и Y коррелированы, то есть связаны линейной зависимостью.

# Заключение

Был проведен статистический анализ данных для переменных X и Y. Были найдены , , , , , , , коэффициенты корреляции X и Y, уравнение линейной регрессии. Так же был построен график, изображающий данные выборки и найденную функцию регрессии. Была проверена гипотеза о значимости выборочного коэффициента корреляции при .

Найденная корреляция стремится к -1, что говорит о зависимости переменных.

# Список литературы

1. Кремер Н. Ш. Теория вероятностей и математическая статистика — Учебник для вузов. — 2-е изд., перераб. и доп. — М.: ЮНИТИ - ДАНА, 2004.
2. Гмурман В. Е. Теория вероятностей и математическая статистика. Учебное пособие для вузов. — М.: Высшая школа, 2001.