ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ

«Международный университет природы, общества и человека «Дубна»

Кафедра высшей математики

КУРСОВАЯ РАБОТА ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

ТЕМА: «ИССЛЕДОВАНИЕ КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ ПО ВЫБОРКЕ»

Вариант №92

Выполнил: студент 2071 группы 2 курса

Серый Сергей Андреевич

Руководители: ст. преп. Чебоненко З.А.

Дубна, 2015

**Оглавление**

Введение

Теоретическая часть

Практическая часть

Заключение

Список используемой литературы

# **Введение**

Математическая статистика - наука о математических методах систематизации и использовании статистических данных для научных и практических выводов. Во многих своих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала, выборки генеральной совокупности.

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.

Целью данной курсовой работы является исследование корреляционной зависимости по выборке.

# **Теоретическая часть**

*Выборочный метод -* статистический метод исследования общих свойств совокупности каких-либо объектов на основе изучения свойств лишь части этих объектов, взятых на выборку.

*Генеральная совокупность, генеральная выборка -* это множество объектов (субъектов), отобранных специальным образом для обследования (опроса). Любые данные, полученные на основании выборочного обследования (опроса), имеют вероятностный характер. На практике это означает, что в ходе исследования определяется не конкретное значение, а интервал, в котором определяемое значение находится.

*Выборочное среднее*  *-* это среднее арифметическое значение признака выборочной совокупности. Если все значения признака выборки объема n различны, то:





*Выборочная дисперсия*  *-* среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения.



*Выборочное средне квадратическое отклонение -* квадратный корень из выборочной дисперсии.



*Парная выборка -* выборка состоящие из одних и тех же объектов обследованных в разные моменты времени.

*Парные выборки* - это две выборки, которые набираются таким образом, что каждым наблюдением одной выборкисопоставлено наблюдение другой выборки сопоставление основывается, как правило, на совпадении значений одной или более заданных характеристик(признаков).

*Выборочная ковариация -* числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений случайных величин от их математических ожиданий.



*Корреляционная зависимость -* статистическая <https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0> взаимосвязь двух или более случайных величин <https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F\_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0> (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми).

*Выборочный коэффициент корреляции -* оценка коэффициента корреляции, служащая для определения линейной связи между величинами X и Y . Коэффициент корреляции изменяется в пределах от -1 до 1.



Выборочный коэффициент корреляции показывает тесноту линейной связи между X и Y - чем ближе к единице абсолютное значение коэффициента, тем сильнее линейная связь между переменными.

*Регрессия* - зависимость среднего значения какой-либо величины *Y* от другой величины *X*. Понятие регрессии в некотором смысле обобщает понятие функциональной зависимости *y = f(x)*. Только в случае регрессии одному и тому же значению *x* в различных случаях соответствуют различные значения *у*.

Регрессионный анализ заключается в определении аналитического выражения связи, в которой изменение одной величины (называемой зависимой или результативным признаком) обусловлено влиянием одной или нескольких независимых величин (факторов).

По форме зависимости различают:

· Линейную регрессию, которая выражается уравнением прямой:



· Нелинейную (параболическую):



*Исследование линейной регрессии:*

Определим коэффициенты линейной функции  методом наименьших квадратов. Для этого составим сумму:



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





Раскрыв скобки, мы получим:





Выразим *a* и *b*:





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**

## корреляция выборка регрессия

## ***Постановка задачи***

· Провести статистический анализ данных для переменных X и Y. Найти выборочные средние, дисперсии и среднеквадратические отклонения для X и Y по отдельности;

· Найти ковариацию Cov (X,Y);

· Найти коэффициент корреляции X и Y;

· Найти по выборке уравнение линейной регрессии (Y как функцию X) по методу наименьших квадратов;

· Найти по выборке уравнение параболической регрессии второго порядка (Y как функцию X) по методу наименьших квадратов;

· Построить графики, изображающие данные выборки и найденные функции регрессии;

· Проверить гипотезу о значимости выборочного коэффициента корреляции при заданном уровне значимости альфа = 0.001;

## **Исходные данные**

Нам дана выборка (объемом *n =* 20*)* зависимости числа *Y* от числа *X*

Таблица 1. Исходные данные. Вариант №92

|  |  |
| --- | --- |
| x | Y |
| 13,3 | 39,1 |
| 18 | 54,7 |
| 20,2 | 59,1 |
| 3,4 | 10,8 |
| 16,5 | 50,9 |
| 5,4 | 24,1 |
| 23,8 | 62,5 |
| 0,5 | 4,6 |
| 18,4 | 56,4 |
| 10,3 | 40,9 |
| 7,6 | 26,1 |
| 0,3 | -0,1 |
| 17,8 | 48,1 |
| 0,4 | 3,9 |
| 15,1 | 52,2 |
| 14,6 | 49 |
| 0,8 | 0 |
| 5,4 | 25,5 |
| 24,6 | 59,7 |
| 8 | 26,5 |

## **Статистический анализ данных для переменных X и Y**

Найдем выборочные средние для X и Y по формуле (1):







Найдем выборочные дисперсии для X и Y по формуле (2) и (3):

















Найдем выборочные среднеквадратические отклонения для X и Y по формуле (4):



7,843634362

21,27653637

## **Ковариация Cov (X, Y)**

Найдем выборочную ковариацию для переменныхX и Yпо формуле(5):





## **Коэффициент корреляции X и Y**

Найдем выборочный коэффициент корреляции для переменных X и Y по формуле (6):





## **Уравнение линейной регрессии**

Найдем коэффициенты *a* и *b* для уравнения линейной регрессии.

Для нахождения коэффициентов *a* и *b* методом наименьших квадратов были посчитаны следующие необходимые параметры:













Тогда коэффициенты  и  соответственно равны:



Таким образом, получаем уравнение линейной регрессии:





Рис. 1 Линейная регрессия

## **Гипотеза о значимости выборочного коэффициента корреляции**

*Проверим гипотезу о значимости выборочного коэффициента корреляции, для этого вычислим*  и  *(на уровне значимости равном 0,001).*





Получаем:



Поскольку , можно сказать что выборочный коэффициент корреляции отличается от нуля, а X и Y коррелированы.

**Заключение**

В данной курсовой работе были вычислены основные числовые характеристики выборок по X и Y.

По выборке проведено исследование корреляционной зависимости по выборке. Найдены ковариация и коэффициент корреляции. В результате проведенной работы были закреплены теоретические знания и приобретены практические навыки работы со статистиками, умение находить точечные и интервальные оценки математического ожидания и дисперсии.

Был изучен метод регрессионного анализа, при помощи которого удаётся провести зависимость прожиточного минимума (Y) от заработной платы (X).

**Список используемой литературы**

1. Кремер, Н.Ш. Теория вероятностей и математическая статистика - Учебник для втузов. - 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.

3. Гмурман, В.Е. «Руководство к решению задач по ТВ и МС.» 2003