# Тема лекционного дистанционного занятия по дисциплине «Статистика» для студентов факультета ПСО (Право и организация социального обеспечения) 3 курс, преподаватель Гасанова С.Р.

# Лекция №10 Корреляционно-регрессионный анализ взаимосвязей

**Понятие корреляции.**Все явления в мире взаимосвязаны. Это значит, что каждое событие оказывает влияние на все события, следующие за ним, а само происходит вследствие всех событий, случившихся до него.

До сих пор рассматривались основные статистические характеристики изолированно друг от друга, теперь будем изучать, как и в какой форме одно явление оказывает влияние на другое. Это является предметом корреляционно-регрессионного анализа.

При статистическом исследовании корреляционных связей одной из основных задач является определение их формы, т.е. построение модели связи.

Построение регрессионной модели проходит несколько этапов:

1. Определение факторов, которые оказывают определяющее воздействие на результативный признак.

2. Определение форм воздействия факторов и результата.

3. Определение степени влияния на результат учтенных и неучтенных факторов.

В статистике изучаются следующие виды связей:

1. Балансовая связь – характеризует зависимость между источниками формирования результатов и их использованием.

2. Компонентные связи – характеризуются тем, что изменение статистического показателя определяется изменением компонентов, входящих в этот показатель, как множители.

3. Факторные связи – характеризуются тем, что они появляются в согласованной вариации изучаемых показателей.

Одни выступают как факторные, другие как результативные.

При функциональной связи изменение результативного признака (у) обусловлено всецело действием одного факторного признака (х), т.е. одному факторному соответствует одно и только одно значение результативного признака y=f(x). Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы изучаемой величины.

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, при большом числе наблюдений, то такая зависимость называется стохастической. Частным случаем стохастической связи является корреляционная, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

Количественно степень тесноты связи определяется при помощи коэффициента корреляции по шкале Чеддока, табл.10.1.

По направлению выделяют связь прямую, т.е. с увеличением или уменьшением значения факторного признака происходит увеличение или уменьшение результата.

Например, увеличение производительности труда способствует увеличению уровня рентабельности.

Таблица 10.1

Шкала Чеддока

|  |  |
| --- | --- |
| Величина коэффициента корреляции | Характер связи |
| до |±0,3| | практически отсутствует |
| |±0,3|-|±0,5| | слабая |
| |±0,5|-|±0,7| | умеренная |
| |±0,7|-|±1,0| | сильная |

И обратную, когда значения результативного признака изменяются под воздействием факторного, но в противоположном направлении.

Например, с увеличением фондоотдачи снижается себестоимость единицы продукции.

По аналитическому выражению выделяют связи прямолинейные и нелинейные.

В статистике не всегда требуются количественные оценки, важно просто определить форму воздействия одних факторов на другие.

Корреляция – это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при котором изменение одной из случайных величин приводит к уменьшению математического ожидания другой. В статистике принято различать следующие варианты зависимостей:

1. Парная корреляция – связь между двумя признаками.

2. Частная корреляция – зависимость между результатом и одним факторным признаком при фиксированном значении других факторных признаков.

3. Множественная корреляция – зависимость результативного и двух или более факторных признаков, включенных в исследование.

Корреляционный анализ имеет своей задачей количественно определить тесноту связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

По направлению связи распределяют:

а) прямую регрессию (положительную);

б) обратную (отрицательную), т.е. с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.

**Методы корреляционно-регрессионного анализа связи показателей.**Наиболее разработанная – метод парной корреляции, рассматривающая влияние вариации факторного признака (х) на результативный (у).

Для выявления связи применяются различные виды уравнения прямолинейной и криволинейной связей. Аналитическая связь между ними может быть описана следующими уравнениями:

Прямая 

Гипербола 

Парабола +а2х2

Определить тип уравнения можно, исследуя зависимость графически. Однако есть более общее указание.

- если результативный и факторный признаки увеличиваются одинаково, примерно в арифметической прогрессии, то связь прямая.

- при обратной – гиперболическая.

- если факторный признак увеличивается в арифметической прогрессии, а результативный быстрее, то парабола или степенная.

Оценка параметров уравнений регрессии а0; а1; а2осуществляется методом наименьших квадратов.

При линейной зависимости, уравнение регрессии выглядит следующим образом



где а0– усредненное влияние на результативный признак случайных факторов. а1– коэффициент регрессии показывает насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

На практике часто исследования проводятся по большому числу наблюдений. В этом случае исходные данные представляют в сводной корреляционной таблице. При этом анализу подвергаются сгруппированные данные и по факторному, и по результативному признаку.

Корреляционная таблица дает общее представление о направлении связи. Если оба признака (х и у) располагаются в возрастающем порядке, а частоты (fxy) сосредоточены по диагонали сверху вниз направо, то связь прямая, в противном случае обратная.



прямая обратная

О тесноте связи между признаками х и у по корреляционной таблице можно судить по кучности расположения частот вокруг диагонали (поскольку заполненные клетки таблицы в стороне от нее).

Если клетки заполнены большими цифрами, то связь слабая. Чем ближе частоты (fxy) располагаются к одной из диагоналей, тем теснее связь. Если в расположении частот (fxy) нет системности, то можно судить об отсутствии связи.

**Измерение тесноты связи.**Измерить тесноту зависимости между величинами (х) и (у) можно при помощи корреляционного отношения () и линейного коэффициента корреляции ().

Первый показатель (корреляционное отношение, или коэффициент корреляции по Пирсону) применим ко всем случаям корреляционной зависимости независимо от формы этой связи.

Второй показатель (линейный коэффициент корреляции), служит мерой тесноты связи лишь для линейной зависимости между (х) и (у).

Следует различать теоретическое корреляционное отношение и эмпирическое.

Теоретическое корреляционное отношение представляет собой относительную величину, получающуюся в результате сравнения среднего квадратического отклонения в ряду выравненных значений результативного признака, т.е. рассчитанных по уравнению регрессии, со средним квадратическим отклонением в ряду эмпирических значений результатов признака. Если первое обозначить через δ, а второе – через σ, то .

Учитывая, что сумма выравненных и эмпирических значений результативного признака совпадает, т.е. , и среднее значение признака у этих рядов одинаково ().

Среднеквадратическое отклонение ряда эмпирических значений результативного признака можно записать как

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-nCM3g6.png | (10.1) |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-oDxoXp.png | (10.2) |

Корреляционное отношение можно записать следующим образом:

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-N9D9iX.png | (10.3) |

Возведя обе части в квадрат, получим

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-EUgBam.png. | (10.4) |

Это корреляционное отношение в квадрате называют коэффициентом детерминации.

Т.е. чем ближе значение корреляционного отношения к 1, тем больше связь у и х. Чем ближе к 0, тем связь слабее. Обычно η меньше 0,3, зависимость маленькая; 0,3- 0,6 – зависимость средняя, больше 0,6 – большая.

**Линейный коэффициент корреляции.**

В случае линейной зависимости между двумя коррелируемыми величинами тесноту связи измеряют линейным коэффициентом корреляции (r), который может быть рассчитан по нескольким формулам:

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-Y7oUAP.png | (10.5) |

где а1- коэффициент регрессии в уравнении связи;

σх- среднее квадратическое отклонение факторного признака;

σу- среднее квадратическое отклонение результативного признака.

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-DTjeMU.png | (10.6) |

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-0XB4K1.png | (10.7) |

Линейный коэффициент корреляции может принимать значения от -1 до +1.

Если r отрицательна – это обратная зависимость между х и у, т.е. с увеличением х уменьшается у и наоборот.

Если r =0 – связь между х и у отсутствует.

Если 0 < r < 1 – связь функциональная.

Следовательно в рассмотренном примере связь между валовой продукцией и основными производственными фондами функциональная.

**Коэффициент эластичности.**В экономическом анализе часто используют рассчитываемые на основе уравнений регрессии коэффициенты эластичности результативного признака относительно факторного.

Коэффициент эластичности (Э) показывает, на сколько процентов изменяется в среднем результативный признак (y) при изменении факторного признака (х) на 1%.

Коэффициент эластичности для большинства форм связи величина переменная, т.е. изменяется с изменением фактора (х).

Так, для линейной зависимости () коэффициент эластичности рассчитывается по формуле

|  |  |
| --- | --- |
| Э=https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-3SJ_ul.png | (10.8) |

**Непараметрические методы оценки корреляционной связи показателей.**При исследовании степени тесноты связи между качественными признаками, каждый из которых представлен в виде альтернативных признаков, возможно использование следующих расчетных таблиц.

|  |  |  |
| --- | --- | --- |
| a | b | a+b |
| c | d | c+d |
| a+c | b+d | a+b+c+d |

Теснота связи между качественными показателями рассчитывается по следующим коэффициентам:

а) коэффициент ассоциации Д.Юла

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-7M9UZ_.png | (10.9) |

Связь имеется если коэффициент больше или равен 0,5.

б) коэффициент контингенции К.Пирсона

|  |  |
| --- | --- |
| https://studfile.net/html/2706/813/html_X8G5ptDHp2.Pl0p/img-Ko4BpD.png | (10.10) |

Связь имеется если коэффициент больше или равен 0,3.