**Лабораторная работа № 5**

**Криволинейная корреляция и регрессия в среде Excel и STATISTICA**

Криволинейная корреляция и регрессия в среде MS Excel

Если связь между изучаемыми явлениями существенно отклоняется от пропорциональной, что легко установить по графику, то коэффициент корреляции непригоден в качестве меры связи. Он может указать на отсутствие сопряженности там, где налицо сильная криволинейная зависимость. Поэтому необходим новый показатель, который правильно измерял бы степень криволинейной зависимости. Таким показателем является корреляционное отношение, обозначаемое греческой буквой *η* (эта). Оно измеряет степень корреляции при любой ее форме.

*Криволинейная связь между признаками* – это такая связь, при которой равномерным изменениям первого признака соответствуют неравномерные изменения второго, причем эта неравномерность имеет определенный закономерный характер.

При графическом изображении криволинейных связей, когда по оси абсцисс откладывают значения первого признака (аргумент – независимая переменная), а по оси ординат – значения второго признака (функция – зависимая переменная) и полученные точки соединяют, получают изогнутые линии. Характер изогнутости зависит от природы коррелируемых признаков.

Корреляционное отношение измеряет степень корреляции при любой ее форме.

В отличие от коэффициента корреляции, который дает одинаковую меру связи признаков (первого со вторым и второго с первым), корреляционное отношение второго признака по первому обычно не бывает равно корреляционному отношению первого признака по второму. Поэтому крайне важно определить какая выборка является аргументом, а какая функцией.

По виду линии на графике можно определить характер связи (прямолинейная или криволинейная), также тип аппроксимации.

Задачей исследователя является подобрать вид функции, которая бы наиболее четко ложилась на поле регрессии, иначе: значение квадрата корреляционного отношения было бы максимально возможным.

***Пример***

Исследовалось влияние вводимого перорально с кормом сорбента (Х – в г) молочным коровам на снижение содержания 137Cs в молоке (Y – в %).

Исходные данные представлены на рисунке 1.



*Рисунок 1 – Исходные данные*

**Шаг 1.** Из панели инструментов выберите и нажмите команду **Мастер диаграмм**. В появившемся окне выберите **тип (Точечная)** и вид диаграммы, как показано на рисунке 2.



*Рисунок 2 – Выбор типа и вида диаграммы*

**Шаг 2**. Нажмите кнопку **Далее**. В окно **Диапазон** установите маркер, выберите в таблице анализируемые выборки (рисунок 3).



*Рисунок 3 – Установка диапазона*

**Шаг 3.** Нажмите кнопку **Далее**. Введите названия осей (рисунок 4).



*Рисунок 4 – Оформление диаграммы*

**Шаг 4.** Щелкните мышкой **Готово**. Результат обработки появится в виде диаграммы.

**Шаг 5.** На графике щелкните правой кнопкой по любой точке диаграммы.

**Шаг 6.** Выберите опцию **Добавить линию тренда** и **Тип**. В меню представлены четыре типа аппроксимации: логарифмическая, полиноминальная, степенная и экспоненциальная. В данном случае полином 2 степени (рисунок 5а).

**Шаг 7**. В опции **Параметры** выберите установки, как показано на рисунке 5б. Щелкнете мышкой **OK**.



а) б)

*Рисунок 5 – Тип аппроксимации и параметры кривой*

Отредактированная диаграмма представлена на рисунке 6.



*Рисунок 6 – Отредактированная диаграмма*

Уравнение регрессии и квадрат корреляционного отношения находятся в правом нижнем углу диаграммы.

Результаты анализа показали, что корреляционная связь между массой вводимого сорбента и уровнем снижения содержания 137Cs в молоке велика r = 0,99 и описывается полиномом 2 степени: **Y = -0,45 x X2 + 11,46 x X - 23,41**

При массе вводимого сорбента больше 12 г наблюдается насыщение. Поэтому оптимальное значение массы сорбента следует рекомендовать в пределах 12–14 г.

**Задание**. Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе для вариантов 1–3 Приложения. Получите результат и сделайте заключение.

Криволинейная корреляция и регрессия в среде STATISTICA

Условия задачи такие же, как в предыдущем разделе лабораторной работы. Ввести исходные данные, как показано на рисунке7.



*Рисунок 7 – Исходные данные Var1 – независимая переменная – X; Var2 – зависимая переменная – Y.*

Проведем анализ в модуле Nonlinear estimation (Нелинейная оценка).

**Шаг 1.** Из **Переключателя модулей** Statistica откройте модуль **Nonlinear estimation (Нелинейная оценка)**. Высветите название модуля и далее щелкните мышью по названию модуля **Nonlinear estimation** (рисунок 8).



*Рисунок 8 – Запуск модуля Nonlinear estimation*

**Шаг 2.** На экране появится стартовая панель модуля. Выберите опцию **User specified regression, least squares (Метод наименьших квадратов)** и далее щелкните мышью по названию модуля (рисунок 9).



*Рисунок 9 – Стартовая панель модуля Nonlinear estimation*

**Шаг 3.** В появившемся окне щелкните мышью по кнопке **Function of estimated** (Предполагаемая функция) (рисунок 11).



*Рисунок 11 – Панель ввода функции*

**Шаг 4**. В окне с клавиатуры введите предполагаемую функцию. В отличие от подобной операции в табличном редакторе MS Excel в Statistica вы можете ввести любую формулу, связывающую зависимую и независимую переменные.

В данном случае предполагается, что наиболее подходящей функцией является полином второй степени типа: или в конкретном случае в соответствии с таблицей исходных данных:  (рисунок 12)

В нижней части рисунка приведен перечень алгебраических и функциональных символов, которые воспринимаются программой.

Нажмите **OK**. Затем еще раз **OK**.



*Рисунок 12 – Ввод функции*

Верхняя часть окна информирует о модели, методе, количестве взятых в анализ пар. В середине окна выберите метод аппроксимации. Например: **Gauss–Newton** (рисунок 13). Нажмите **OK.**

**Шаг 5.** В верхней части появившегося окна результатов (рисунок 14) показаны значения корреляционного отношения и его квадрата, 0,99 и 0,98. Это указывает на сильную корреляционную связь между переменными.



*Рисунок 13 – Панель пуска аппроксимации*



*Рисунок 14 – Окно результатов*

**Шаг 6.** В окне результатов щелкните мышью по кнопке **Summary: Parameters & standard errors (Итоговые параметры и стандартные ошибки)**. Полученные результаты (рисунок 15) подкрашены красным цветом, что свидетельствует о достоверности аппроксимации функцией:



В столбце **Estimate (Оценка)** показаны значения коэффициентов: a, b, c. Далее указаны стандартные ошибки, t–критерий при 22 степенях свободы, уровень значимости меньше 0,05, верхний и нижний пределы достоверности.



*Рисунок 15 – Результаты аппроксимации*

**Шаг 7**. Щелкните мышью по кнопке **Fitted 2D function & observed vals (Подогнанная функция)**. На рисунке 16 вы увидите графическую интерпретацию корреляционной связи исходных массивов в виде заданной функции: 



*Рисунок 16 – Графическая интерпретация корреляционной связи*

**Шаг 8**. В окне результатов (рисунок 14) в режиме Quick нажмите кнопку Analysis of Variance (Анализ вариантов). Результат выполненной операции представлен на рисунке 17, который свидетельствует о достоверности регрессии (F = 8806,2 при р < 0,00..).



*Рисунок 17 – Результат анализа вариантов*

**Шаг 9**. В окне рисунка 14 перейдите в режим просмотра результатов **Residuals (Остатки)**. Щелкните мышью по кнопке **Observed, predicted, residual vals (Наблюдаемый, предсказанный, остаточный)**. Результаты выполненной операции представлены на рисунке 18.



*Рисунок 18 – Наблюдаемые и аппроксимированные значения функции*

**Шаг 10.** Для оценки адекватности модели щелкните мышкой по кнопке **Observed vs. Predicted (Наблюдаемые против предсказанных)** (рисунок 19).

Из рисунка видно, массивы наблюдаемых и предсказанных значений описываются линейной функцией . При этом k = 1 и коэффициент парной корреляции близок к 1.



*Рисунок 19 –* Визуализация результатов анализа

К заключению, сделанному в предыдущем разделе, следует добавить, что корреляция и регрессия достоверны, так как F = 8806,2 (рисунок 17) и ta = 7,8, tb = 17,9 и tc =14,1 (рисунок 15), что существенно выше критических значений при р < 0,00..

**Задание**. Выполните расчетные процедуры в соответствии с порядком операций, выполненных в настоящем разделе для вариантов 4–7 Приложения. Получите результат и сделайте заключение.

**Приложение**

Задания к лабораторным работам

|  |
| --- |
| Варианты заданий |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| X | Y | X | Y | X | Y | X | Y | X | Y | X | Y | X | Y |
| 14,5 | 49,8 | 33,3 | 0,03 | 9,06 | 2,13 | 0,9 | 66,9 | 0,97 | 0,4 | 1,5 | 0,7 | 0,87 | 0,15 |
| 5,1 | 22,5 | 7,1 | 0,2 | 8,95 | 2,12 | 0,01 | 0,01 | 1,21 | 2,79 | 1,47 | 0,69 | 0,98 | 0,11 |
| 6,1 | 28,2 | 11,2 | 0,13 | 5,58 | 1,61 | -0,8 | -46,4 | 0,83 | 0,09 | 1,58 | 0,73 | 1,32 | 0,04 |
| 6,5 | 31,2 | 14,3 | 0,08 | 11,26 | 2,38 | 1,2 | 165,1 | 1,16 | 1,83 | 1,57 | 0,73 | 1,22 | 0,05 |
| 6,7 | 33,5 | 17,2 | 0,06 | 9,14 | 2,14 | 1,5 | 314,8 | 0,95 | 0,31 | 1,45 | 0,68 | 0,81 | 0,18 |
| 8,2 | 41,8 | 22,2 | 0,05 | 2,87 | 0,83 | 1,1 | 141,1 | 1,53 | 22,9 | 1,45 | 0,69 | 1,47 | 0,02 |
| 7,4 | 34,9 | 34,7 | 0,03 | 12,01 | 2,42 | -0,1 | -0,1 | 1,09 | 1,13 | 1,02 | 0,51 | 0,96 | 0,11 |
| 10,7 | 46,9 | 11,2 | 0,1 | 6,86 | 1,81 | 0,4 | 5,8 | 0,48 | 0 | 1,18 | 0,58 | 0,77 | 0,2 |
| 14,3 | 48,8 | 3,7 | 0,37 | 7 | 1,89 | 0,8 | 60,5 | 1,1 | 1,15 | 0,92 | 0,46 | 0,93 | 0,12 |
| 8,1 | 40,3 | 33,8 | 0,03 | 2,86 | 0,99 | -0,1 | -0,3 | 1,15 | 1,73 | 0,96 | 0,48 | 0,88 | 0,14 |
| 8,1 | 40,3 | 10,9 | 0,12 | 7,59 | 1,97 | 1,2 | 175 | 1,04 | 0,73 | 1,78 | 0,79 | 0,96 | 0,11 |
| 4,6 | 21,3 | 13,1 | 0,08 | 10,47 | 2,25 | 2,6 | 1845,6 | 1,23 | 3,15 | 0,55 | 0,2 | 0,73 | 0,23 |
| 8,4 | 42,2 | 18,1 | 0,07 | 6,78 | 1,77 | 1,6 | 421,1 | 0,75 | 0,04 | 1,21 | 0,6 | 1,56 | 0,02 |
| 9,6 | 44,9 | 20,1 | 0,06 | 2,57 | 0,69 | 1,2 | 192,9 | 1,02 | 0,62 | 0,86 | 0,43 | 1,25 | 0,05 |
| 8,8 | 43,9 | 23,4 | 0,05 | 7,59 | 1,96 | 0,7 | 42,4 | 1,11 | 1,3 | 0,61 | 0,26 | 1,27 | 0,04 |
| 9,1 | 44,3 | 30,1 | 0,04 | 9,59 | 2,17 | -0,01 | -0,01 | 1,14 | 1,63 | 1,09 | 0,54 | 0,68 | 0,26 |
| 9,6 | 44,8 | 27 | 0,04 | 11,08 | 2,35 | 0,02 | 0,02 | 1,35 | 7,22 | 0,86 | 0,43 | 1,12 | 0,07 |
| 7,4 | 36,2 | 25,4 | 0,04 | 12,33 | 2,43 | 0,2 | 1,1 | 1,11 | 1,31 | 1,12 | 0,56 | 1,53 | 0,02 |
| 8,5 | 43,9 | 25,6 | 0,04 | 11,76 | 2,42 | 1,6 | 443,8 | 1,78 | 88,7 | 0,53 | 0,19 | 0,72 | 0,23 |
| 11,2 | 47,1 | 30,1 | 0,04 | 9,49 | 2,2 | 0,4 | 8,4 | 0,79 | 0,06 | 1,16 | 0,58 | 1,13 | 0,07 |
| 12,7 | 47,9 | 27,7 | 0,04 | 12,2 | 2,48 | -0,3 | -2,8 | 0,84 | 0,1 | 1,05 | 0,52 | 1,12 | 0,07 |
| 13,4 | 48,2 | 26 | 0,05 | 9,62 | 2,22 | -0,5 | -13,1 | 0,83 | 0,1 | 0,57 | 0,22 | 0,94 | 0,12 |
| 13,8 | 48,7 | 23,2 | 0,05 | 12,53 | 2,48 | 0,8 | 50 | 1,21 | 2,76 | 0,9 | 0,45 | 1,32 | 0,04 |
| 7,2 | 34,1 | 19 | 0,06 | 3,51 | 1,06 | 0,3 | 4,3 | 1,2 | 2,61 | 0,75 | 0,36 | 0,95 | 0,12 |
| 8 | 40,3 | 32,1 | 0,03 | 4,76 | 1,46 | 0,6 | 19,8 | 1,31 | 5,75 | 0,86 | 0,43 | 1,2 | 0,06 |
|  |  | 16,8 | 0,07 | 10,41 | 2,3 | 0,2 | 1,1 | 1,67 | 49,21 | 0,77 | 0,37 | 0,72 | 0,23 |
|  |  | 12,1 | 0,09 | 8,71 | 2,09 | 0,6 | 23,2 | 1,18 | 2,29 | 1,61 | 0,74 | 0,69 | 0,25 |
|  |  | 9,3 | 0,17 | 8,7 | 2,12 | 1,4 | 255,9 | 0,65 | 0,01 | 0,68 | 0,31 | 0,96 | 0,11 |
|  |  | 6,7 | 0,26 | 9,68 | 2,17 | 0,8 | 44,8 | 0,89 | 0,17 | 1,01 | 0,51 | 0,9 | 0,14 |
|  |  | 21,5 | 0,05 | 10,04 | 2,24 | 1,5 | 326,4 | 0,68 | 0,01 | 1,36 | 0,65 | 1,67 | 0,01 |