**Лабораторная работа № 3**

**Корреляционный анализ при помощи Excel и STATISTICA**

Корреляционный анализ при помощи Excel

Исходные данные представлены в таблице (в табличном редакторе EXCEL данные представлены двумя столбцами):



Открыть модуль **Анализ данных** выбрать опцию **Корреляция**, после чего щелкнуть мышкой **OK**. В появившемся окне выполнить операции и установки, как показано на рис.1. Щелкнуть мышкой **OK**.



*Рис.1 – Стартовая панель*

Результат обработки появится в указанном поле (выходной интервал $Е$1):



**Задание 1.** Надо было установить, есть ли корреляция между высотой головы *х* и длиной 3-го членика усика у *Drosophila funebris*. Для этого с помощью окуляр микрометра получены следующие данные по *х* и *у* (в делениях окуляр-микрометра):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 15 | 16 | 15 | 15 | 16 | 16 | 17 | 18 | 18 | 17 | 17 | 17 | 15 | 16 | 15 | 15 | 15 | 17 |
| у | 29 | 31 | 32 | 33 | 32 | 33 | 33 | 36 | 36 | 35 | 35 | 35 | 35 | 33 | 31 | 31 | 31 | 35 |
| x | 15 | 13 | 15 | 14 | 17 | 15 | 16 | 15 | 15 | 16 | 15 | 16 | 15 | 16 | 18 | 17 | 14 | 15 |
| у | 33 | 30 | 32 | 31 | 35 | 33 | 33 | 32 | 30 | 33 | 33 | 33 | 30 | 31 | 34 | 34 | 31 | 33 |
| x | 14 | 15 | 15 | 13 | 15 | 16 | 14 | 15 | 15 | 15 | 14 | 15 | 15 | 15 | 16 | 18 | 15 | 14 |
| у | 31 | 31 | 33 | 30 | 30 | 33 | 30 | 33 | 31 | 32 | 30 | 31 | 31 | 32 | 33 | 35 | 32 | 32 |

Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте вывод.

**Задание 2.** Получены следующие данные о продолжительности беременности у кро­ликов породы шиншилла при различных размерах помета (число крольчат в помете *х* и длительность беременности в днях *у)*:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 1 | 8 | 3 | 5 | 7 | 8 | 4 | 8 | 3 | 4 | 4 | 8 | 8 | 5 | 7 | 6 | 6 | 5 | 6 | 6 |
| у | 33 | 30 | 31 | 31 | 31 | 32 | 31 | 31 | 32 | 33 | 32 | 31 | 31 | 31 | 31 | 30 | 31 | 32 | 32 | 31 |
| x | 6 | 5 | 7 | 8 | 10 | 6 | 7 | 6 | 7 | 6 | 5 | 10 | 7 | 8 | 8 | 6 | 5 | 6 | 5 | 4 |
| у | 32 | 32 | 31 | 32 | 31 | 31 | 30 | 31 | 31 | 32 | 31 | 30 | 32 | 32 | 31 | 31 | 31 | 32 | 30 | 31 |
| x | 6 | 8 | 6 | 5 | 8 | 7 | 6 | 5 | 9 | 5 | 3 | 4 | 7 | 8 | 9 | 5 | 6 | 2 | 2 | 4 |
| у | 31 | 31 | 32 | 32 | 31 | 30 | 32 | 31 | 31 | 31 | 32 | 32 | 31 | 31 | 31 | 31 | 31 | 32 | 33 | 33 |

Есть ли корреляция между длительностью плодоношения и размерами помета? Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте вывод.

**Задание 3.** Были получены следующие данные о весе ягнят-баранчиков (одинцов) *у* и весе баранов – их отцов – *х* (в кг).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 76,6 | 72,2 | 67,0 | 66,5 | 63,3 | 65,4 | 63,9 | 63,1 | 63,0 | 62,5 | 62,2 |
| у | 4,56 | 4,79 | 4,49 | 4,32 | 4,59 | 4,32 | 4,67 | 4,29 | 4,57 | 4,20 | 4,12 |
| x | 61,0 | 60,2 | 60,0 | 59,6 | 59,5 | 58,9 | 58,0 | 57,8 | 57,6 | 57,0 |  |
| у | 4,13 | 4,70 | 3,80 | 4,23 | 3,76 | 4,08 | 4,61 | 4,37 | 4,30 | 4,0 |  |
| x | 56,8 | 55,4 | 55,0 | 53,8 | 53,7 | 52,0 | 51,4 | 51,0 | 50,9 | 48,5 |  |
| у | 3,82 | 4,12 | 4,19 | 4,16 | 4,09 | 4,12 | 4,02 | 4,31 | 4,06 | 4,03 |  |

Есть ли корреляция между весом баранчиков и весом их отцов? Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте вывод.

Корреляционный анализ в STATISTICA

Массив исходных данных приведен в таблице ранее (для табличного редактора EXCEL). Перенесите его в STATISTICA (рис. 2).



*Рис. 2 – Исходные данные*

**X** – независимая переменная

**Y** – зависимая переменная

Проведите анализ в модуле **Basic statistics/Tables (основная статистика)**. Рассмотрим и установим связь между X и Y.

**Шаг 1**. Из **Переключателя модулей** STATISTICA откройте модуль **Basic statistics/Tables (основная статистика)**. Высветите название модуля и далее щелкните мышью по названию модуля: **Basic statistics/Tables** (рис. 3).



*Рис. 3 – Стартовая панель модуля Basic statistics/Tables*

**Шаг 2.** На экране появится (рис. 4). Щелкните мышью по названию **Correlation matrics (корреляционная матрица)**.



*Рис. 4*

**Шаг 3**. Выберите переменные для анализа. Выбор переменных осуществляется с помощью кнопки **Two lists**, находящейся в центре верхней части панели (рис. 5).



*Рис. 5 – Стартовая панель модуля Correlation matrics*

После того как кнопка будет нажата, диалоговое окно **Select one or two variable list (выбрать списки зависимых и независимых переменных)** появится на вашем экране (рис. 6).



*Рис. 6 – Окно выбора переменных для анализа*

**Шаг 4**. Высветив имя переменной в правой части окна, выберите переменную в левой части окна. После нажатия кнопки **OK** в режиме **Options** выполните установки, показанные на рис. 7, подсветив **Displey detaled table of results**.



Рис. 7

**Шаг 5**. После нажатия кнопки **Summary** программа произведет расчеты корреляции межу X и Y, и через секунду на экране появится следующее окно результатов (рис.7):



*Рис. 8 – Результат расчета корреляции*

− среднее;

− стандартное отклонение;

− значение коэффициента корреляции r;

− значение коэффициента детерминации r2 ;

− t – критерий;

− р – уровень значимости;

− число коррелируемых пар;

− свободный член – 12,65659.

− коэффициент при независимой переменной – 0.816349.

В этом примере r = 0,98… Это очень хорошее значение (подсвечено красным цветом), показывающее, что построенная регрессия объясняет более 90% разброса значений переменной X относительно среднего.

Из полученной таблицы видно, что оцененная модель имеет вид:

**Y = 12,657 + 0,81635 \* X**

**Шаг 6**. После нажатия кнопки **2D scatterplots** появится график, на котором данные с подогнанной прямой имеют вид (рис. 9).



*Рис. 9 – Линейная регрессия для данных X и Y*

*Интерпретация значения корреляций*

Коэффициент корреляции Пирсона (*r*) представляет собой меру линейной зависимости 2 переменных. Если возвести его в квадрат, то полученное значение коэффициента детерминации (r2) представляет долю вариации, общую для 2 переменных (иными словами, степень зависимости или связанности двух переменных). Чтобы оценить зависимость между переменными, нужно знать как величину корреляции, так и ее значимость.

Уровень значимости, вычисленный для каждой корреляции, представляет собой главный источник информации о надежности корреляции. Значимость определенного коэффициента корреляции зависит от объема выборок. Критерий значимости основывается на предположении, что распределение остатков (т.е. отклонений наблюдений от регрессионной прямой) для зависимой переменной *y* является нормальным (с постоянной дисперсией для всех значений независимой переменной *x*).

*Выбросы*. По определению, выбросы являются нетипичными, резко выделяющимися наблюдениями. Так как при построении прямой регрессии используется сумма квадратов расстояний наблюдаемых точек до прямой, то выбросы могут существенно повлиять на наклон прямой и, следовательно, на значение коэффициента корреляции. Поэтому единичный выброс (значение которого возводится в квадрат) способен существенно изменить наклон прямой и, следовательно, значение корреляции.

Заметим, что если размер выборки относительно мал, то добавление или исключение некоторых данных (которые, возможно, не являются выбросами) способно оказать существенное влияние на прямую регрессии (и коэффициент корреляции).

Обычно считается, что выбросы представляют собой случайную ошибку, которую следует контролировать. К сожалению, не существует общепринятого метода автоматического удаления выбросов. Чтобы не быть введенными в заблуждение полученными значениями, необходимо проверить на диаграмме рассеяния каждый важный случай значимой корреляции. Очевидно, выбросы могут не только искусственно увеличить значение коэффициента корреляции, но также реально уменьшить существующую корреляцию.

Некоторые исследователи применяют численные методы удаления выбросов. Например, исключаются значения, которые выходят за границы ±2 стандартных отклонений (и даже ±1,5 стандартных отклонений) вокруг выборочного среднего. К сожалению, в общем случае, определение выбросов субъективно, и решение должно приниматься индивидуально в каждом эксперименте (с учетом особенностей эксперимента или сложившейся практики в данной области).

**Задание 4**. У 100 серебристо-черных лисиц были измерены (в см) длина туловища *х* и длина хвоста *у*:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 70 | 65 | 66 | 65 | 71 | 68 | 64 | 57 | 66 | 65 | 67 | 62 | 67 | 62 | 63 | 57 | 64 | 66 | 69 | 58 |
| у | 40 | 40 | 40 | 40 | 40 | 42 | 39 | 38 | 41 | 43 | 39 | 45 | 43 | 38 | 40 | 40 | 41 | 45 | 43 | 37 |
| x | 63 | 67 | 67 | 67 | 65 | 65 | 67 | 70 | 65 | 71 | 69 | 64 | 64 | 66 | 69 | 72 | 66 | 66 | 67 | 66 |
| у | 45 | 38 | 39 | 37 | 42 | 38 | 38 | 38 | 38 | 40 | 39 | 43 | 43 | 42 | 40 | 41 | 47 | 47 | 40 | 40 |

 Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте вывод.

**Задание 5.** Были получены следующие данные о весе *х* (в г) левой камеры сердца и длине ядер *у* (в *μ*) в мышцах сердца:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| х | 207 | 221 | 256 | 262 | 273 | 289 | 291 | 292 | 304 | 328 | 372 | 397 | 460 | 632 |
| У | 16,6 | 18,0 | 15,9 | 20,7 | 19,4 | 19,8 | 11,7 | 21,0 | 23,0 | 13,6 | 19,6 | 22,9 | 19,4 | 28,4 |

Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте обоснованный вывод.

**Задание 6.** Были получены следующие данные о весе *х* (в кг) и длине туловища *у* (в см) 100 серебристо-черных лисиц:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 4,7 | 4,6 | 5,2 | 5,1 | 5,3 | 5,3 | 4,6 | 4,8 | 5,8 | 5,7 | 4,5 | 5,7 | 5,0 | 4,8 | 4,7 | 5,2 | 4,6 |
| y | 70 | 65 | 69 | 70 | 66 | 68 | 65 | 71 | 69 | 68 | 57 | 73 | 65 | 67 | 71 | 62 | 69 |
| x | 5,5 | 5,5 | 4,8 | 4,7 | 6,0 | 5,1 | 5, | 4,5 | 5,0 | 5,0 | 4,9 | 5,5 | 5,2 | 5,6 | 5,2 | 5,7 | 5,3 |
| y | 62 | 63 | 67 | 64 | 64 | 66 | 68 | 69 | 58 | 63 | 67 | 74 | 67 | 67 | 70 | 65 | 71 |
| x | 5,4 | 5,3 | 4,6 | 5,6 | 5,1 | 4,9 | 5,2 | 5,3 | 5,0 | 5,3 | 5,6 | 5,0 | 5,1 | 5,5 | 5,6 | 5,2 | 5,0 |
| y | 63 | 64 | 64 | 66 | 63 | 69 | 62 | 72 | 66 | 66 | 67 | 67 | 66 | 63 | 67 | 62 | 71 |
| x | 5,5 | 5,6 | 5,0 | 6,7 | 4,7 | 5,3 | 5,0 | 5,1 | 5,0 | 5,1 | 4,8 | 5,0 | 6,0 | 5,5 | 4,6 | 4,5 | 4,5 |
| y | 67 | 66 | 66 | 69 | 64 | 69 | 70 | 62 | 68 | 68 | 72 | 68 | 67 | 66 | 69 | 65 | 65 |
| x | 5,4 | 5,0 | 4,9 | 5,0 | 5,7 | 5,9 | 5,6 | 5,1 | 5,1 | 4,6 | 4,9 | 6,2 | 5,6 | 5,2 | 5,1 | 4,5 |  |
| y | 65 | 65 | 64 | 66 | 66 | 67 | 62 | 63 | 64 | 69 | 69 | 68 | 65 | 69 | 67 | 68 |  |
| x | 4,8 | 5,5 | 6,0 | 5,3 | 4,8 | 5,3 | 5,1 | 5,4 | 4,7 | 5,0 | 5,9 | 5,0 | 5,2 | 5,6 | 5,2 | 5,1 |  |
| y | 61 | 64 | 62 | 66 | 59 | 65 | 62 | 68 | 61 | 67 | 69 | 69 | 66 | 66 | 67 | 70 |  |

Определите, есть ли корреляция между весом и длиной туловища у лисиц? Определите коэффициент корреляции и постройте линию регрессии, зарисуйте ее в рабочую тетрадь. Сделайте обоснованный вывод.