**Задание 9 Простая регрессия**

**Задача 1.**

Для исследования на основе простой регрессии воспользуемся файлом данных exam.sav. Он содержит 10-балльную оценку нервной возбудимости (тревожности) 36 студентов и количество решенных ими зачетных тестовых задний (из 20 возможных). Гипотеза о линейности отношения этих двух переменных говорит о том, что чем выше нервная возбудимость студента, тем выше его результативность (например, потому, что спокойных студентов меньше волнуют их знания, а тревожные студенты проводят больше времени за подготовкой к зачету). В качестве зависимой переменной (критерия) выступает переменная *тест,* а в качестве независимой переменной (предиктора) — переменная *трев.*

Необходимо прогнозировать значения переменной *тест* по известным значениям переменной *трев,* используя уравнение регрессии.

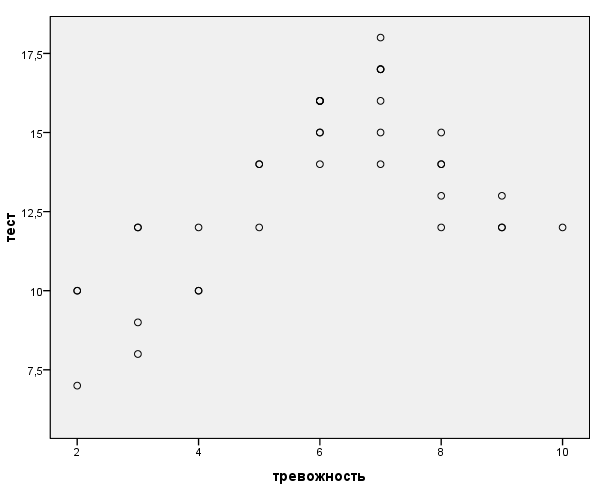
**Оценка криволинейности**

Выдвинутая гипотеза о том, что увеличение нервной возбудимости перед экзаменом всегда улучшает результат студента, очевидно, вызывает сомнения. Более разумными кажутся следующие соображения. При низкой возбудимости результаты экзамена должны быть низкими, поскольку излишнее спокойствие снижает потребность студента в подготовке. С увеличением возбудимости результат до определенного момента должен улучшаться, однако слишком возбудимые студенты вряд ли способны сконцентрироваться и показать хороший результат.

Поэтому наилучших показателей должны добиваться те студенты, чей уровень возбудимости является промежуточным. При регрессионном анализе соотношение между зависимой и независимой переменными считается линейным. Однако возможная ошибка прогноза может быть велика. Можно предположить, что если изменить вид общего уравнения (например, включить в него квадрат переменной *трев*), прогнозируемые значения будут ближе к реальным.

Сначала необходимо построить диаграмму рассеяния. Для этого в меню выберите **Графика**>**scatterplot**

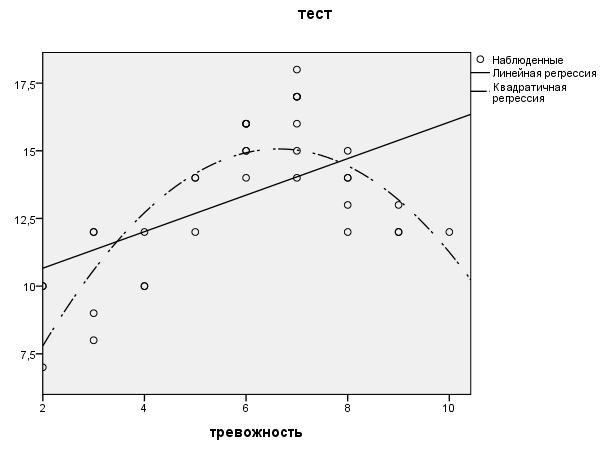
По вертикальной оси диаграммы отложены значения переменной *тест*, по горизонтальной — значения переменной *трев.* Очевидно, что зависимость между переменными не является линейной: у нее имеется выраженный максимум, а при движении к краям диаграммы наблюдается убывание значений переменной *тест*.



Для того чтобы статистически оценить криволинейность, в меню **Анализ**, подменю **Регрессия** предусмотрена команда **Подгонка кривых**.

В ее диалоговом окне необходимо задать зависимую переменную (*тест*), независимую переменную (*трев*) и установить флажки **Линейная** и **Квадратичная**.

После щелчка на кнопке OK появятся диаграмма, демонстрирующая линейную и криволинейную тенденции:





По диаграмме можно оценить, насколько близка к линейной или квадратичной зависимость между значениями переменных.

В таблице результатов включены значения коэффициентов *B* регрессии (Константа b0, b1, b2), поэтому можно составить линейное и квадратичное уравнения регрессии для прогнозируемых значений.

Константа и коэффициенты уравнений указаны в последних трех столбцах выводимых результатов.

А уравнение для квадратичной регрессии выглядит следующим образом:

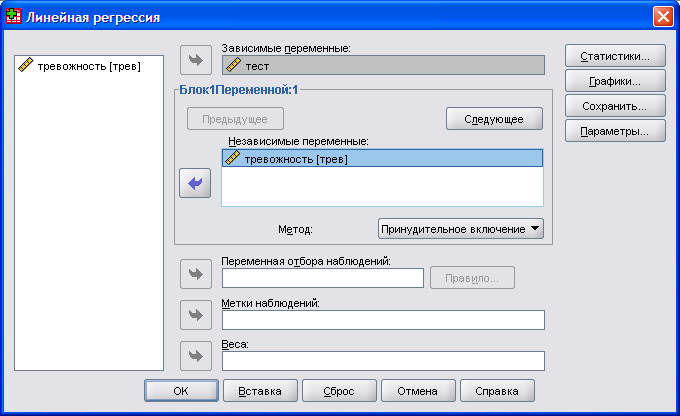
тестпрогноз = 0,1615 + 4,4896 × трев – 0,3381 × (трев)2

В случае линейной регрессии величина *R2* (столбец R квадрат в таблице выводимых результатов) равна 0,298, то есть 29,8 % дисперсии переменной *тест* обусловлено воздействием со стороны переменной *трев*.

В то же время для квадратичной регрессии, которая учитывает и линейную, и криволинейную связи, *R2* = 0,675, то есть она обусловливает 67,5 % дисперсии переменной *тест*. Малый *p*-уровень для обоих уравнений свидетельствует об очень высокой статистической достоверности полученных результатов.

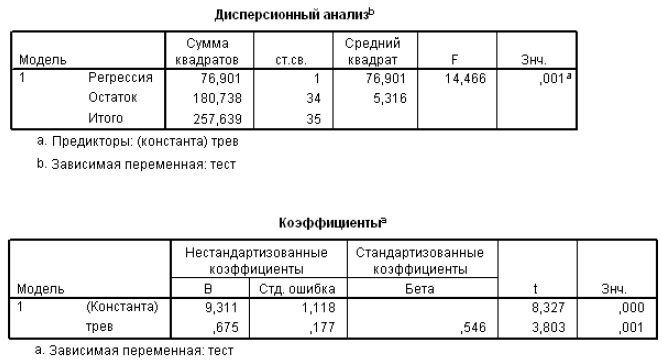
Очевидно, что квадратичная регрессия описывает отношения между переменными *тест* и *трев* более адекватно, чем линейная.

В меню **Анализ** выберите команду **Регрессия**>**Линейная.** В диалоговом окне нужно выбрать Зависимую и Независимую переменные и щелкнуть на кнопке OK.



После нажатия на «ОК», в окне вывода мы получаем три таблицы с данными описывающими модель:





Для линейной регрессии уравнение имеет вид:

тестпрогноз = 9,311 + 0,6751 × трев

**Терминология, используемая при выводе**

**R** — поскольку в анализе участвовала единственная независимая переменная, эта величина равна коэффициенту корреляции (r) между переменными тест и трев.

**R квадрат** — квадрат величины R (R2), равный доле дисперсии переменной тест, обусловленной воздействием переменной трев.

**Скорректированный R квадрат** — скорректированная величина R2. Величина R2, используемая в расчетах, на практике оказывается несколько завышенной. Скорректированная величина R2 менее формальна и ближе к реальным результатам.

**Стд. ошибка оценки** — в таблице Сводка для модели это стандартное отклонение оценок значений зависимой переменной тест.

**Регрессия** — статистики, оценивающие долю дисперсии зависимой переменной, обусловленную влиянием независимых переменных.

**Остаток** — статистики, оценивающие долю дисперсии зависимой переменной, не обусловленную влиянием независимых переменных.

**ст. св.** — число степеней свободы, для регрессии оно равно числу независимых переменных. Для остатка — равно разности размера выборки и числа степеней свободы регрессии, уменьшенной на единицу (36 – 1 – 1 = 34).

**Знч.** — величина р-уровня значимости, вероятность случайности полученного результата.

**B** — коэффициент и константа b0 линейного уравнения регрессии:

**Стд. ошибка** — стандартная ошибка, в таблице Коэффициенты это характеристика стабильности коэффициента B, равная стандартному отклонению коэффициентов B, рассчитанных для большого числа выборок из генеральной совокупности.

**Бета** — стандартизованный коэффициент регрессии (β). Представляет собой коэффициент B для стандартизованных значений переменной трев. Для линейных отношений эта величина всегда лежит в диапазоне от –1,0 до 1,0, а для криволинейных отношений может выходить за границы этого диапазона.

**t** — отношение коэффициента B к его стандартной ошибке.

**Сумма квадратов, Средний квадрат, F** — показатели, относящиеся к дисперсионному анализу.

При описании результатов необходимо предоставить

диаграмму рассеяния,

уравнение регрессии

оценку модели (см. лекцию 7).

Задача 2. **Анализ криволинейных зависимостей**

При составлении квадратного уравнения регрессии SPSS вычисляет новую переменную, значения которой равны квадратам соответствующих значений переменной трев.

Квадратичная зависимость графически имеет вид параболы — кривой с двумя ветвями, которые могут быть направлены вверх, если коэффициент при квадратном члене положительный, или вниз, если коэффициент при квадратном члене отрицательный. Поскольку квадратное уравнение имеет не только квадратный, но и положительный линейный член, изображаемый в виде возрастающей наклонной прямой, конец левой ветви параболы расположен ниже, чем конец правой ветви.

Для того чтобы включить в анализ регрессии квадрат переменной трев, нам необходимо «вручную» создать переменную трев2 с помощью команды меню **Преобразовать > Вычислить переменную** содержащую квадраты значений переменной трев.

В меню Анализ выберите команду Регрессия>Линейная регрессия, чтобы открыть диалоговое окно Линейная регрессия. Если вы уже работали с этим окном, щелкните на кнопке Сброс.

Щелкните сначала на переменной тест, чтобы выделить ее, а затем — на верхней кнопке со стрелкой, чтобы переместить переменную в поле Зависимая переменная.

Щелкните сначала на переменной трев, чтобы выделить ее, а затем — на второй сверху кнопке со стрелкой, чтобы переместить переменную в список Независимые переменные.

Повторите предыдущее действие для переменной трев2.

Щелкните на кнопке OK, чтобы открыть окно вывода.

Опишите полученные данные так же, как в задаче 1.

**Задача 3**

1. Используя исходные данные, построить модель парной линейной регрессии для изучения зависимости *х* от *у*.

2. Найти коэффициент корреляции и коэффициент детерминации.

3. Сделать прогноз показателя при конкретном значении.

4. Найти коэффициент эластичности, т.е. на какую величину в среднем изменится показатель *Y* при изменении *X* на единицу.

При описании результатов необходимо предоставить

диаграмму рассеяния,

уравнение регрессии

оценку модели (см. лекцию 11).

**Варианты заданий**

Номер варианта – это ваш номер в списке группы по журналу. При несовпадении работа не допускается к проверке.

**Вариант 1**

*х* - энерговооруженность (кВт) на десяти предприятиях,

*у -* производительность труда (тыс. руб.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 2,8 | 2,2 | 3,0 | 3,5 | 3,2 | 3,7 | 4,0 | 4,8 | 6,0 | 5,4 |
| *Y* | 6,7 | 6,9 | 7,2 | 7,3 | 8,4 | 8,8 | 9,1 | 9,8 | 10,6 | 10,7 |

Сделать прогноз показателя *Y* при *х*=5.

**Вариант 2**

*х* - качество земли (балл),

*у -* урожайность (ц/га)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 32 | 33 | 35 | 37 | 38 | 39 | 40 | 41 | 42 | 44 |
| *Y* | 19,5 | 19,0 | 20,5 | 21,0 | 20,8 | 21,4 | 23,0 | 23,3 | 24,0 | 24,5 |

3. Сделать прогноз показателя *Y* при *х*=43.

**Вариант 3**

*х* - качество земли (балл),

*у -* урожайность (ц/га)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 45 | 46 | 47 | 49 | 50 | 52 | 54 | 55 | 58 | 60 |
| *Y* | 24,2 | 25,0 | 27,0 | 26,8 | 27,2 | 28,0 | 30,0 | 30,2 | 32,0 | 33,0 |

3. Сделать прогноз показателя *Y* при *х*=51.

**Вариант 4**

*х* – товарооборот,

*у-* издержки обращения по отношению к товарообороту

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | *7* | 10 | 15 | 20 | 30 | 45 | 60 | 120 |
| *Y* | 10,0 | 9,0 | 7,5 | 6,0 | 6,3 | 5,8 | 5,4 | 5,0 |

3. Сделать прогноз показателя *Y* при *х*=16.

**Вариант 5**

*х* - электровооруженность на одного рабочего,

*у -* выпуск готовой продукции на одного рабочего

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 2 | 5 | 3 | 7 | 2 | 6 | 4 | 9 | 8 | 4 |
| *Y* | 3 | 6 | 4 | 6 | 4 | 8 | 6 | 9 | 9 | 5 |

3. Сделать прогноз показателя Y при х=10.

**Вариант 6**

х - качество земли (балл),

у - урожайность (ц/га).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 35 | 37 | 38 | 39 | 40 | 41 | 42 | 44 |
| *Y* | 23,0 | 23,3 | 24,0 | 24,5 | 24,2 | 25,0 | 27,0 | 28,0 |

3. Сделать прогноз показателя Y при х=45.

**Вариант 7**

*х* - производительность труда;

*у -* прибыль.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 |
| *Y* | 2,6 | 2,4 | 3,3 | 2,9 | 3,7 | 4,2 | 5,5 | 6,4 |

3. Сделать прогноз показателя *Y* при *х*=0,9.

**Вариант 8**

х -производительность труда ;

у - рентабельность производства.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 0,9 | 1,5 | 2,0 | 2,5 | 2,8 | 3,0 | 1,2 | 1,4 |
| *Y* | 3,1 | 5,1 | 5,9 | 6,1 | 7,2 | 8,1 | 3,8 | 5,3 |

3. Сделать прогноз показателя Y при х=3,1.

**Вариант 9**

х - среднемесячная оплата труда (ден. ед.);

у- уровень рентабельности торговой деятельности.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 127 | 133 | 149 | 156 | 162 | 182 | 197 | 201 |
| *Y* | 2,01 | 2,12 | 2,77 | 3,62 | 3,80 | 4,01 | 4,33 | 4,5 |

3. Сделать прогноз показателя Y при х=150.

**Вариант 10**

*х* - среднемесячный товарооборот (на душу населения);

у - уровень рентабельности

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 21 | 22 | 27 | 28 | 29 | 33 | 35 | 38 |
| *Y* | 2,01 | 2,12 | 3,62 | 2,77 | 3,80 | 4,01 | 4,33 | 4,6 |

3. Сделать прогноз показателя Y при х=25.

**Вариант 11**

*х -* удельный вес продовольственных товаров в товарообороте;

у - уровень рентабельности

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 67 | 70 | 73 | 74 | 77 | 83 | 84 | 85 |
| *Y* | 4,33 | 4,01 | 3,80 | 3,62 | 2,77 | 2,01 | 2,12 | 2,0 |

3. Сделать прогноз показателя Y при х=75.

**Вариант 12**

х - средняя взвешенная цена акций;

у - объем продаж

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 84,4 | 82,4 | 80,1 | 63,4 | 76,2 | 75,1 | 74,8 | 73,0 |
| *Y* | 79,5 | 279 | 71,4 | 243 | 76,3 | 74,7 | 211 | 75 |

3. Сделать прогноз показателя Y при х=81.

**Вариант 13**

В таблице представлены статистические данные о размере товарооборота X и суммы издержек обращения Y по девяти магазинам

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Товарооборот, X | 480 | 510 | 530 | 540 | 570 | 590 | 640 | 650 | 660 |
| Издержки обращения, Y | 19 | 25 | 30 | 28 | 29 | 32 | 36 | 36 | 37 |

3. Сделать прогноз показателя Y при х=500.

**Вариант 14**

Обозначим через X цену оптовой продажи некоторого товара, через Y – цену его розничной продажи.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 68 | 70 | 70,5 | 71 | 72 | 73 | 74 | 76 | 76 | 77 | 79 | 80 |
| Y | 79 | 80 | 81 | 80,5 | 82 | 81 | 82 | 82 | 82,5 | 83,5 | 84 | 85 |

3. Сделать прогноз показателя Y при х=70.

**Вариант 15**

Таблица содержит данные о росте X и массе Y для выбранных наугад студентов.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 163 | 165 | 165 | 170 | 173 | 175 | 178 | 180 | 183 | 185 | 188 | 193 |
| Y | 68 | 66 | 61 | 66 | 75 | 63 | 67 | 84 | 75 | 75 | 82 | 90 |

3. Сделать прогноз показателя Y при х=176.

**Вариант 16**

Данные о годовой производительности труда в расчете на одного рабочего Y и энерговооруженности труда Х на предприятиях од­ной отрасли

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 6,7 | 6,9 | 7,2 | 7,3 | 8,4 | 8,8 | 9,1 | 9,8 | 10,6 | 10,7 | 11,1 | 11,8 | 12,1 | 12,4 |
| Y | 2,8 | 2,2 | 3,0 | 3,5 | 3,2 | 3,7 | 4,0 | 4,8 | 6,0 | 5,4 | 5,2 | 5,4 | 6,0 | 9,0 |

3. Сделать прогноз показателя Y при х=10.

**Вариант 17**

Случайная выборка 10 фармацевтических фирм показала следующее соотношение между прибылью Y и затратами на научные исследования X.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 20 | 40 | 40 | 30 | 10 | 40 | 50 | 30 | 20 | 50 |
| Y | 50 | 50 | 60 | 40 | 10 | 30 | 50 | 50 | 30 | 60 |

3. Сделать прогноз показателя Y при х=45.

**Вариант 18**

На 10 территориях были измерены процентный показатель перенаселенности X и показатель детской смертности Y.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 13 | 33 | 12 | 40 | 12 | 7 | 20 | 4 | 15 | 26 |
| Y | 124 | 151 | 124 | 156 | 128 | 78 | 127 | 104 | 127 | 144 |

3. Сделать прогноз показателя Y при х=10.

**Вариант 19**

Фондоотдача оборудования X и удельный вес продукции высшей категории качества Y.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1,47 | 1,25 | 1,82 | 1,45 | 1,75 | 1,37 | 1,61 | 1,93 | 1,68 | 1,66 |
| Y | 34,08 | 35,89 | 36,93 | 32,31 | 34,91 | 30,20 | 31,23 | 48,13 | 30,08 | 42,86 |

3. Сделать прогноз показателя Y при х=1,5.

**Вариант 20**

В таблице содержатся данные, показывающие связь между количеством дней X, проведенных пациентами в больнице, и затратами больницы Y, которые компенсируются страховой компанией.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 3 | 6 | 7 | 2 | 4 | 12 | 15 | 5 | 9 |
| Y | 50 | 175 | 180 | 200 | 60 | 140 | 420 | 540 | 170 | 300 |

3. Сделать прогноз показателя Y при х=10.

**Вариант 21**

Крупная корпорация проверяет соотношение между прибылью Y и процентом используемых производственных мощностей Х для каждого из 12 заводов, входящих в корпорацию.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 50 | 57 | 61 | 68 | 77 | 80 | 82 | 85 | 89 | 91 | 95 | 99 |
| Y | 2,5 | 6,9 | 3 ,1 | 4,6 | 7,3 | 45 | 6 1 | 11 6 | 100 | 14,2 | 161 | 195 |

3. Сделать прогноз показателя Y при х=70.

**Вариант 22**

В таблице приводится связь между валовым национальным продуктом (ВНП) на душу населения X и процентным показателем Y грамотности взрослого населения для 22 стран и территорий в середине 1970-х гг.

|  |  |  |
| --- | --- | --- |
| Страна | X | Y |
| Непал | 45 | 5,0 |
| Бирма | 57 | 47,5 |
| Уганда | 64 | 27,5 |
| Южный Вьетнам | 76 | 17,5 |
| Таиланд | 96 | 68,0 |
| Гаити | 105 | 10,5 |
| Индонезия | 131 | 17,5 |
| Южн. Корея | 144 | 77,0 |
| Гана | 172 | 22,5 |
| Перу | 179 | 47,5 |
| Сальвадор | 219 | 39,4 |
| Британская Гвиана | 235 | 74,0 |
| Гонконг | 272 | 57,5 |
| Панама | 329 | 65,7 |
| Ливан | 362 | 47,5 |
| Сингапур | 400 | 50,0 |
| Аргентина | 490 | 86,4 |
| Исландия | 572 | 98,5 |
| Чехословакия | 680 | 97,5 |
| Франция | 943 | 96,4 |
| Новая Зеландия | 1310 | 98,5 |
| Канада | 1947 | 97,5 |

3. Сделать прогноз показателя Y при х=1000.