## Лабораторная работа 3. НЕЛИНЕЙНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

|  |
| --- |
| КРАТКАЯ СПРАВКА |
| Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.  *Нелинейные регрессии* делятся на два класса:  -        регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и  -        регрессии, нелинейные по оцениваемым параметрам.  Регрессии, *нелинейные по объясняющим переменным*:  -        полиномы разных степеней ;  -        равносторонняя гипербола .  Регрессии, *нелинейные по оцениваемым параметрам:*  -        степенная ;  -        показательная ;  -        экспоненциальная .  Нелинейная регрессия по включенным переменным определяется, как и в линейной регрессии, методом наименьших квадратов (МНК), ибо эти функции линейны по параметрам.  Так, в *параболе второй степени* , заменяя переменные , , получим двухфакторное уравнение линейной регрессии:  .  Соответственно, для *полинома  k-го порядка*  .  получим линейную модель множественной регрессии с *k* объясняющими переменными:  .  В уравнении *равносторонней гиперболы* –  делаем замену *z=1/x,* получаем линейное уравнение  .  Построению моделей регрессий,нелинейных по оцениваемым параметрам, предшествует процедура линеаризации.  Для *степенной модели*  линеаризация производится путём логарифмирования обеих частей уравнения  .  С помощью замены  получаем линейное уравнение .  Для *показательной модели*  линеаризация производится также с помощью логарифмирования обеих частей уравнения . С помощью замены   получаем линейное уравнение  .  Тесноту связи изучаемых явлений для нелинейной регрессии оценивает индекс корреляции :  .  Величина данного показателя находится в границах: , чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии.  Оценку качества построенной модели даст коэффициент (индекс) детерминации, а также средняя ошибка аппроксимации.  Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака *y* характеризует *коэффициент (индекс) детерминации* . *Коэффициент детерминации* – квадрат индекса корреляции.  *Средняя ошибка аппроксимации* – среднее отклонение расчетных значений от фактических:  .  Допустимый предел значений  – не более 8-10%.  Средний коэффициент эластичности  показывает, на сколько процентов в среднем по совокупности изменится результат *y* от своей средней величины при изменении фактора *x* на 1% от своего среднего значения:  .  Проверка существенности в целом уравнения нелинейной регрессии осуществляется с помощью  *F-критерия Фишера*. Для этого выполняется сравнение фактического  и критического (табличного)  значений *F-критерия Фишера*.  определяется  из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:  ,  где *n* – число единиц совокупности;  *m* – число параметров при переменных *x*. |

## РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

### *Пример 3.1.* По территориям Центрального и Волго-Вятского районов известны данные за ноябрь 1997 г. (табл. 3.1).

Таблица 3.1

|  |  |  |
| --- | --- | --- |
| Район | Средняя заработная плата и выплаты социального характера, тыс. руб., *y* | Прожиточный минимум в среднем на душу населения, тыс. руб., *x* |
| *Центральный* |  |  |
| Брянская обл. | 615 | 289 |
| Владимирская обл. | 727 | 338 |
| Ивановская обл. | 584 | 287 |
| Калужская обл. | 753 | 324 |
| Костромская обл. | 707 | 307 |
| Орловская обл. | 657 | 304 |
| Рязанская обл. | 654 | 307 |
| Смоленская обл. | 693 | 290 |
| Тверская обл. | 704 | 314 |
| Тульская обл. | 780 | 304 |
| Ярославская обл. | 600 | 341 |
| *Волго-Вятский* |  |  |
| Респ. Марий Эл | 554 | 364 |
| Респ. Мордовия | 560 | 342 |
| Чувашская Респ. | 545 | 310 |
| Кировская обл. | 672 | 411 |
| Нижегородская обл. | 796 | 304 |

***Требуется.***

1.     Постройте поле корреляции и сформируйте гипотезу о форме связи.

2.     Рассчитайте параметры уравнений степенной, экспоненциальной и гиперболической парной регрессии.

3.     Оцените тесноту связи с помощью показателей корреляции и детерминации.

4.     Дайте с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

5.     Оцените качество уравнений с помощью средней ошибки аппроксимации.

6.     Оцените статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7.     Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости a = 0,05.

8.     Оцените полученные результаты.

***Решение.***

1. Создайте рабочую книгу *ПРИМЕР3.xlsx*. Для каждого вида регрессии используйте новый лист.

На листе рабочей книги *ПРИМЕР3.xlsx* постройте таблицу 3.1, несколько изменив ее внешний вид (рис. 3.1).

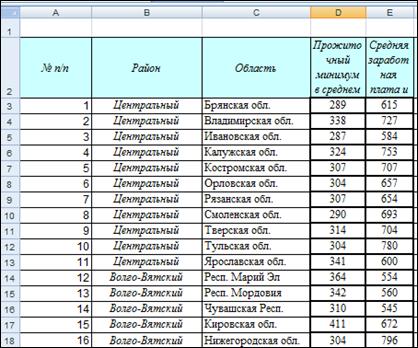


Рис. 3.1. Исходные данные

Используя мастер диаграмм, постройте точечную диаграмму зависимости средней заработной платы от прожиточного минимума (рис. 3.2).



Рис. 3.2. Поле корреляции

*По полю корреляции можно предположить, что корреляционная связь между признаками практически отсутствует*.

2. ***Степенная регрессия*** имеет вид: 

Линеаризуем данную модель. В примере линеаризация производится путем логарифмирования обоих частей уравнения.

Введем новые переменные: .

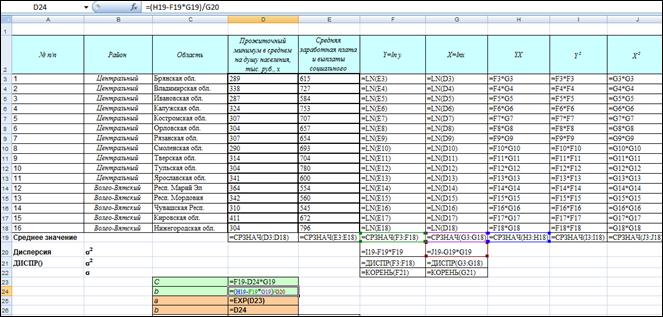
В результате получим линейную регрессионную модель:

,  где  .

Коэффициенты данного уравнения регрессии вычисляются по формулам:

,  .

Для расчетов используем данные таблицы на рисунке 3.3.



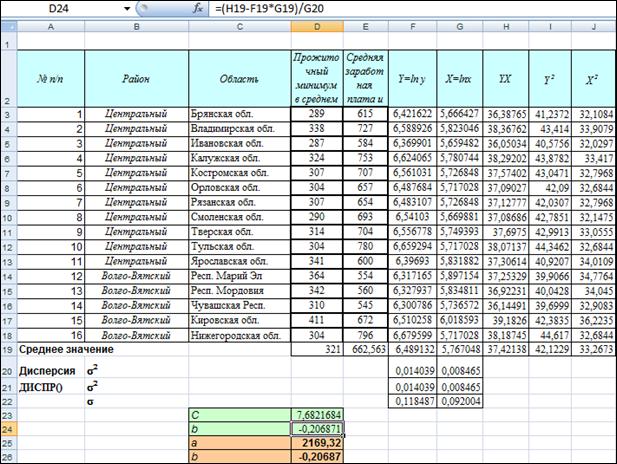


Рис. 3.3. Расчет параметров уравнения степенной регрессии

Получим:  

Отсюда пересчитаем исковые коэффициенты модели:

; .

*В результате эмпирическое уравнение степенной регрессии имеет вид: .*

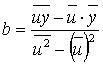
***Экспоненциальная регрессия*** имеет вид: . Множитель *0,005* введен для удобства расчетов.

Линеаризуем данную модель. Для этого введем новую переменную: .

В результате получим линейную регрессионную модель:

.

Коэффициенты данного уравнения регрессии вычисляются по формулам:

, .

Для расчетов используем данные таблицы на рисунке 3.4.

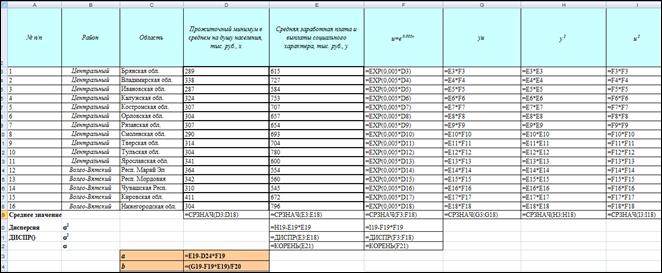




Рис. 3.4. Расчет параметров уравнения экспоненциальной регрессии

Получим: ; .

*В результате эмпирическое уравнение экспоненциальной регрессии имеет вид: .*

***Гиперболическая регрессия*** имеет вид: . Множитель 1000 взят для удобства расчетов.

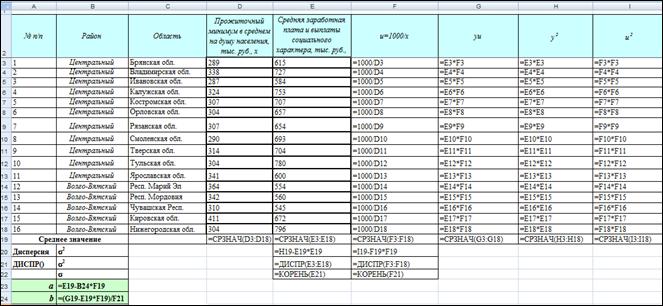
Уравнение равносторонней гиперболы линеаризуется при замене: .

В результате получим линейную регрессионную модель:.

Коэффициенты данного уравнения регрессии вычисляются по формулам:

, .

Для расчетов используем данные таблицы на рисунке 3.5.



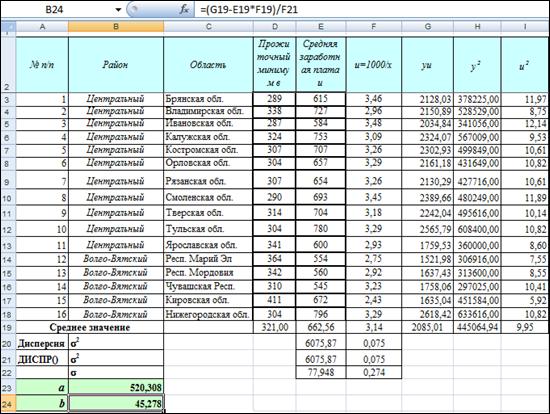


Рис. 3.5. Расчет параметров уравнения гиперболической регрессии

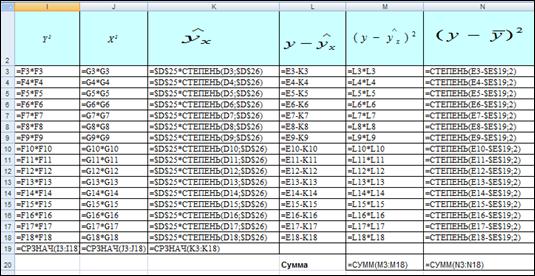
Получим: ; 

*В результате эмпирическое уравнение гиперболической регрессии имеет вид: .*

3. Коэффициент детерминации вычисляется по формуле: .

Определим коэффициенты детерминации для разных моделей регрессии. Для этого для всех моделей добавим столбцы, вычисляющие .

Например, для ***парной степенной регрессии*** (рис. 3.6) (**):



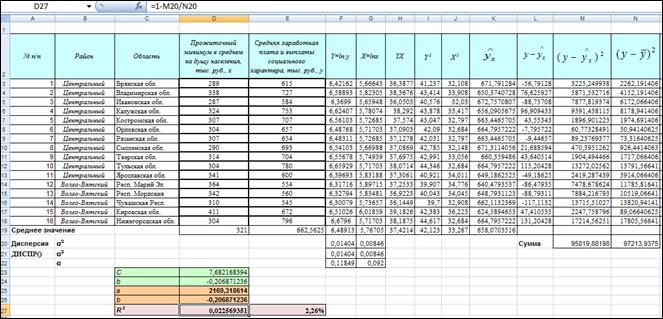
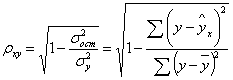


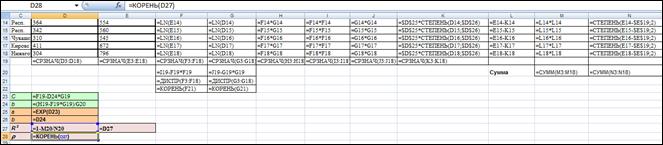
Рис. 3.6. Определение коэффициента детерминации для парной степенной регрессии

Для парной степенной регрессии: . Величина коэффициента  показывает, что доля дисперсии, объясняемая регрессией, в общей дисперсии результативного признака *y* составляет 2,26%.

Тесноту связи изучаемых явлений для нелинейной регрессии оценивает индекс корреляции :

.

Для парной степенной регрессии (рис. 3.7) индекс корреляции . На основании шкалы Чеддока теснота связи слабая.



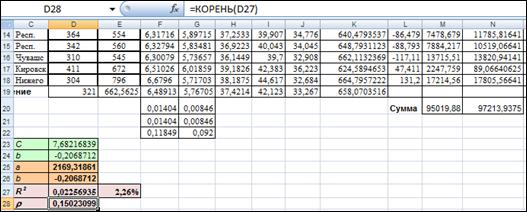


Рис. 3.7. Определение индекс корреляции для парной степенной регрессии

Аналогично рассчитывается коэффициент детерминации и индекс корреляции для других моделей:

-       для парной экспоненциальной регрессии:

.  - связь прямая слабая;

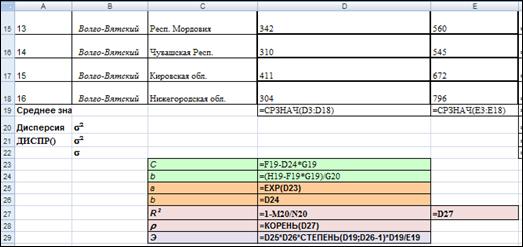
-       для парной гиперболической регрессии:

. - связь прямая слабая.

4. Средний коэффициент эластичности равен: 

Т.к. , , то для ***степенной модели*** (рис. 3.8) имеем:

, *т.е. при изменении прожиточного минимума в среднем на душу населения на 1% от своего среднего значения средняя заработная плата и выплаты социального характера уменьшатся в среднем на 0,2052% .*



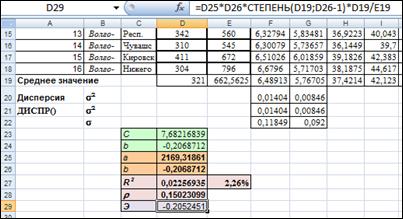


Рис. 3.8. Определение среднего коэффициента эластичности для степенной модели

Аналогично

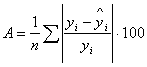
-       для ***экспоненциальной модели:***

;

-       для ***гиперболической модели:***

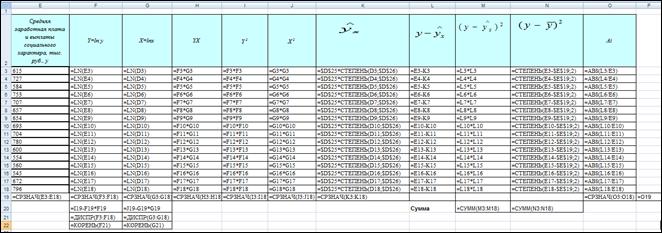
.

5. Средняя ошибка аппроксимации вычисляется по формуле:

.

Определим теперь среднюю ошибку аппроксимации для разных моделей регрессии. Для всех моделей регрессии добавьте столбец, вычисляющий .

Например, для ***парной степенной регрессии*** имеем (рис. 3.9).



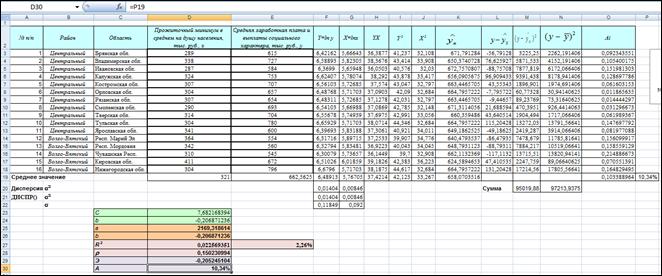


Рис. 3.9. Определение средней ошибки аппроксимации для степенной модели

Аналогично

-       для парной степенной регрессии: ;

-       для парной экспоненциальной регрессии: ;

-       для парной гиперболической регрессии: .

*Для всех моделей средняя ошибка аппроксимации превышает допустимый предел в 10%.*

6. Определим расчетное значение *F*-статистики через найденные коэффициенты детерминации, используя соотношение:

.

Отсюда расчетное значение *F*-статистики равно:

для степенной модели  (рис. 3.10);



Рис. 3.10. Определение *F*-статистики для степенной модели

для экспоненциальной модели ;

для гиперболической модели .

*Все модели по F-критерию Фишера не значимы при доверительной вероятности 95%. Действительно, критическое значение F-статистики (=FРАСПОБР(0,05;1;16-2)) равно:*

*.*

*Расчетные значения F-статистик всех моделей оказались значительно меньше критического уровня. Следовательно, принимается гипотеза H0 и признается статистическая незначимость, ненадежность рассмотренных нелинейных уравнений регрессии.*

7. Полученные результаты, в целом неудовлетворительные. Между средней заработной платой и прожиточным минимумом по данной выборке данных зависимость не обнаружена.

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

***Задача 3.1.*** По территориям Уральского и Западно-Сибирского районов известны данные за ноябрь 1997 г.

|  |  |  |
| --- | --- | --- |
| Область | Потребительские расходы на душу населения, тыс. руб., *y* | Средняя заработная плата и выплаты социального характера, тыс. руб., *x* |
| Республика Башкортостан | 461 | 912 |
| Удмуртская Республика | 524 | 809 |
| Курганская обл. | 298 | 748 |
| Оренбургская обл. | 351 | 847 |
| Пермская обл. | 624 | 1087 |
| Свердловская обл. | 584 | 1074 |
| Челябинская обл. | 425 | 1008 |
| Республика Алтай | 277 | 682 |
| Алтайский край | 321 | 697 |
| Кемеровская обл. | 573 | 1251 |
| Новосибирская обл. | 576 | 967 |
| Омская обл. | 588 | 898 |
| Томская обл. | 497 | 1263 |
| Тюменская обл. | 863 | 3027 |

ТРЕБУЕТСЯ:

1.     Постройте поле корреляции и сформируйте гипотезу о форме связи.

2.     Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, обратной и гиперболической парной регрессии.

3.     Оцените тесноту связи с помощью показателей корреляции и детерминации.

4.     Дайте сравнительную оценку силы связи фактора с результатом с помощью коэффициента эластичности.

5.     Оцените качество уравнений с помощью средней ошибки аппроксимации.

6.     Оцените статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7.     Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости a=0,05.

8.     Оцените полученные результаты.

***Задача 3.2.*** По территориям Центрального района известны данные за 1995 г.

|  |  |  |
| --- | --- | --- |
| Район | Средний размер назначенных ежемесячных пенсий, тыс. руб., *y* | Прожиточный минимум в среднем на одного пенсионера в месяц, тыс. руб., *x* |
| Брянская обл. | 240 | 178 |
| Владимирская обл. | 226 | 202 |
| Ивановская обл. | 221 | 197 |
| Калужская обл. | 226 | 201 |
| Костромская обл. | 220 | 189 |
| г. Москва | 250 | 302 |
| Московская обл. | 237 | 215 |
| Орловская обл. | 232 | 166 |
| Рязанская обл. | 215 | 199 |
| Смоленская обл. | 220 | 180 |
| Тверская обл. | 222 | 181 |
| Тульская обл. | 231 | 186 |
| Ярославская обл. | 229 | 250 |

ТРЕБУЕТСЯ:

1.     Постройте поле корреляции и сформируйте гипотезу о форме связи.

2.     Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, обратной и гиперболической парной регрессии.

3.     Оцените тесноту связи с помощью показателей корреляции и детерминации.

4.     Дайте с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

5.     Оцените качество уравнений с помощью средней ошибки аппроксимации.

6.     Оцените статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7.     Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 10% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости a=0,05.

8.     Оцените полученные результаты.