**Задача по эконометрике**

**Требуется:**

1. Сформулируйте гипотезу о направлении связи.
2. Постройте поле корреляции.
3. Оцените тесноту связи с помощью показателя ковариации и коэффициента корреляции.
4. Рассчитайте параметры уравнения линейной регрессии (Y от Х).
5. Оцените общее качество полученного уравнения.
6. Рассчитайте *F*- статистику.
7. Проверьте значимость коэффициентов уравнения.
8. Проверьте гипотезу об отсутствии автокорреляции.
9. Оформите вывод.

**Вариант 1**

|  |  |  |
| --- | --- | --- |
| Номер региона | Среднедушевой прожиточный минимум в день одного трудоспособного, руб.,  | Среднедневная заработная плата, руб.,  |
| 1 | 81 | 124 |
| 2 | 77 | 131 |
| 3 | 85 | 146 |
| 4 | 79 | 139 |
| 5 | 93 | 143 |
| 6 | 100 | 159 |
| 7 | 72 | 135 |
| 8 | 90 | 152 |
| 9 | 71 | 127 |
| 10 | 89 | 154 |
| 11 | 82 | 127 |
| 12 | 111 | 162 |

 **Это другая решенная задача (как надо делать)**

**Пример**. В таблице приведены данные за 8 лет о совокупном доходе  и расходах на конечное потребление .

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 10 | 12 | 11 | 12 | 14 | 15 | 17 | 20 |
|  | 7 | 8 | 8 | 10 | 11 | 12 | 14 | 16 |

**Требуется:**

1. Сформулировать гипотезу о направлении связи.
2. Построить поле корреляции.
3. Оценить тесноту связи с помощью показателя ковариации и коэффициента корреляции.
4. Рассчитать параметры уравнения линейной регрессии (Y от Х).
5. Оценить общее качество полученного уравнения.
6. Рассчитать *F*- статистику.
7. Проверить значимость коэффициентов уравнения.
8. Проверить гипотезу об отсутствии автокорреляции.
9. Оформить вывод.

Решение.

**1.** Изучение взаимосвязи между переменными начинается с выдвижения гипотезы о наличии и направлении связи. В нашем примере можно предположить, что расходы на конечное потребление положительно связаны с совокупным доходом (т.е. с увеличение дохода расходы также увеличиваются).

Проверим наше предположение.

**2.** *Графический метод.*

Заметим вначале, что объем выборки . Представим ряды значений графически.

Из рисунка видно, что точки  сосредоточены в области, очерченной вытянутым эллипсом, поэтому можно предположить наличие сильной положительной линейной связи между показателями. Т.о. анализ графика подтвердил нашу гипотезу.



**3.** Для определения направления и тесноты взаимосвязи с помощью *статистических методов* рассчитаем показатель ковариации и коэффициент линейной корреляции. Для сокращения расчетов будем постепенно формировать таблицу.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| 1 | 10 | 7 | -3,875 | -3,75 | 14,53125 | 15,01563 | 14,0625 |
| 2 | 12 | 8 | -1,875 | -2,75 | 5,15625 | 3,515625 | 7,5625 |
| 3 | 11 | 8 | -2,875 | -2,75 | 7,90625 | 8,265625 | 7,5625 |
| 4 | 12 | 10 | -1,875 | -0,75 | 1,40625 | 3,515625 | 0,5625 |
| 5 | 14 | 11 | 0,125 | 0,25 | 0,03125 | 0,015625 | 0,0625 |
| 6 | 15 | 12 | 1,125 | 1,25 | 1,40625 | 1,265625 | 1,5625 |
| 7 | 17 | 14 | 3,125 | 3,25 | 10,15625 | 9,765625 | 10,5625 |
| 8 | 20 | 16 | 6,125 | 5,25 | 32,15625 | 37,51563 | 27,5625 |
| **Σ** | **111** | **86** | **-** | **-** | **72,75** | **78,875** | **69,5** |

Вычислим средние арифметические значения обоих признаков:

, .

Найдем отклонения от средних арифметических **** и **** и занесем их в 4-й и 5-й столбцы таблицы.

Умножим  на **** и сложим полученные произведения (последняя строка в 6-м столбце).

Возведем отклонения в квадрат и сложим (последняя строка в 7-м и 8-м столбцах).

Тогда показатель ковариации: .

. Это подтверждает гипотезу о наличии положительной связи между показателями.

Найдем средние квадратические отклонения, пользуясь формулой:

, .

Исходя из формулы коэффициента линейной корреляции, получим:

.

Полученный коэффициент корреляции близок к +1, это указывает на сильную положительную линейную связь между расходами на конечное потребление и совокупным доходом.

Проверим значимость полученного коэффициента корреляции.

Для этого рассчитаем - статистику: .

Зададим уровень значимости коэффициента корреляции (вероятность ошибки): . По формуле находим число степеней свободы:  .

Из таблицы критических значений распределения Стьюдента следует: .

Т.к.  (12,958>3,707), следовательно, мы можем отвергнуть нулевую гипотезу о равенстве теоретического коэффициента корреляции нулю, и, как следствие, гипотезу об отсутствии связи между показателями. Т.е. связь между показателями, выраженная коэффициентом корреляции, статистически значима с вероятностью 0,99 (риск ошибки 0,01 или 1%).

Итогом проведённых расчётов может стать вывод о том, что расходы на конечное потребление действительно линейно положительно связны с величиной располагаемого дохода, и связь эта является статистически значимой. Т.о. проверяемая нами гипотеза получила свое подтверждение на выбранном массиве данных.

**4.** Т.к. расходы на конечное потребление действительно линейно положительно связны с величиной располагаемого дохода, то можно построить уравнение линейной регрессии.

Для составления уравнения регрессии находим коэффициенты  и :

; .

Тогда *уравнение регрессии* принимает вид: .

Проинтерпретируем построенное уравнение регрессии.

В нашем примере коэффициент регрессии  показывает, на какую величину изменятся расходы на конечное потребление, если совокупный доход возрастёт на одну единицу.

Свободный член уравнения  определяет прогнозируемое значение расходов на конечное потребление при величине совокупного дохода, равной нулю.

**5.** Анализ общего качества.

Самым простым способом оценить качество полученного уравнения является *графический способ*: Строим линию регрессии на корреляционном поле. Для этого берем любые две точки значения признака , удобные для вычислений:

 

По данным точкам проводим линию регрессии.



Анализ графика показывает, что уравнение достаточно точно описывает исследуемую зависимость, т.е. разброс точек, представляющих исходные данные, вокруг линии регрессии невелик.

Для анализа общего качества оценённой зависимости *на количественном уровне* используют коэффициент детерминации  (который в случае парной регрессии равен коэффициенту линейной корреляции).

Формула коэффициента детерминации с поправкой на число степеней свободы имеет вид:

,

где – дисперсия, характеризующая общий разброс;

– остаточная дисперсия (среднеквадратическая ошибка).

Для расчёта коэффициента детерминации сначала необходимо найти ошибку регрессии  для каждого года исследуемого периода: , где теоретические значения зависимой переменной *y*, которые рассчитываются по уравнению регрессии .

Продолжим формировать таблицу:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **1** | **2** | **3** | **9** | **10** | **11** |
| 1 | 10 | 7 | 7,1755 | -0,1755 | 0,0308 |
| 2 | 12 | 8 | 9,0201 | -1,0201 | 1,040604 |
| 3 | 11 | 8 | 8,0978 | -0,0978 | 0,009565 |
| 4 | 12 | 10 | 9,0201 | 0,9799 | 0,960204 |
| 5 | 14 | 11 | 10,8647 | 0,1353 | 0,018306 |
| 6 | 15 | 12 | 11,787 | 0,213 | 0,045369 |
| 7 | 17 | 14 | 13,6316 | 0,3684 | 0,135719 |
| 8 | 20 | 16 | 16,3985 | -0,3985 | 0,158802 |
| **Σ** | **111** | **86** |  - |  - | 2,399369 |

Среднеквадратическая ошибка: .

Выборочная исправленная дисперсия зависимой переменной:

.

Тогда значение скорректированного коэффициента детерминации:

.

Замечание. Коэффициент без поправки равен: .

Значение рассчитанного коэффициента детерминации близко к 1, это свидетельствует о достаточно высоком качестве построенного уравнения. Около 96 % разброса зависимой переменной объясняется с помощью данного уравнения.

**6.** Оценка значимости уравнения регрессии в целом**.**

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включённых в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Оценка значимости уравнения проводится с помощью *F*- критерия Фишера.

Величина *F* – критерия связана с коэффициентом детерминации и рассчитывается по формуле: .

*F* – критерий служит для проверки нулевой гипотезы  о том, что все коэффициенты регрессии, за исключением свободного члена *а*, равны нулю и, следовательно, фактор *х* не оказывает влияния на результат *y*  или .

Находим табличное критическое значение *F*-критерия, зададим уровень значимости 0,01: .

Т.к. , то можно отклонить и сделать вывод о существенности статистической связи между *y* и *x.*

**7.** Проверим значимость коэффициентов уравнения зависимости расходов на конечное потребление от располагаемого дохода.

Рассчитаем стандартную ошибку коэффициента регрессии , которая определяется по формуле:

,

где  – остаточная дисперсия на одну степень свободы.

Стандартная ошибка параметра определяется по формуле:

.

Далее рассчитаем *t –* статистики:

,

.

Они служат для проверки нулевых гипотез о том, что истинное значение коэффициента регрессии *b* или свободного члена *a* равно нулю:

.

Альтернативная гипотеза имеет вид: .

*t –* статистики имеют *t –* распределение Стьюдента с (*n-*2) степенями свободы. По таблицам распределения Стьюдента при выбранном уровне значимости α=0,01 и (*n-2*)=6 степенях свободы находят критическое значение .

Т.к. , то нулевая гипотеза должна быть отклонена, коэффициент *b* является статистически значимым.

Т.к. , то нулевая гипотеза не может быть отклонена. Коэффициент *а* статистически незначим.

Т.о. в итоге проведения тестов можно сделать вывод о значимости отличия от нуля оценки коэффициента регрессии *b*, одновременно можно предложить исключить свободный член *а* из уравнения регрессии, т.е. рекомендуется оценить новое уравнение в виде , где  . (Новое уравнение имеет вид .

**8.** Проверка предпосылок МНК. Автокорреляция ошибок регрессии.

Следующим этапом оценивания качества построенного уравнения является проверка предпосылок, лежащих в основе метода расчёта параметров МНК. Для проверки каждой из предпосылок применения МНК имеются специальные тесты. Реализация многих из этих тестов предполагает значительный объём исходных данных, поэтому мы проверим четвёртую предпосылку – отсутствие автокорреляции первого порядка ошибок регрессии.

Эта проверка может быть осуществлена двумя способами:

1. *Графический*.

Заключается в построении графика зависимости остатков (ошибок регрессии) от времени и визуальном определении наличия или отсутствия автокорреляции.

1. *Использование статистики Дарбина-Уотсона DW*.

А) Построим график поведения ошибок регрессии во времени (на основании столбца 10):



Анализ графика не позволяет выявить закономерность поведения ошибок. Т.е. можно предположить, что ошибки распределены независимо (автокорреляция отсутствует).

Б) Эта проверка проводится с помощью расчёта и анализа значения коэффициента Дарбина-Уотсона: .

Для расчётов продолжим формировать таблицу:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | **10** | **11** | **12** |
| 1 | -0,1755 | 0,0308 | - |
| 2 | -1,0201 | 1,040604 | 0,713349 |
| 3 | -0,0978 | 0,009565 | 0,850637 |
| 4 | 0,9799 | 0,960204 | 1,161437 |
| 5 | 0,1353 | 0,018306 | 0,713349 |
| 6 | 0,213 | 0,045369 | 0,006037 |
| 7 | 0,3684 | 0,135719 | 0,024149 |
| 8 | -0,3985 | 0,158802 | 0,588136 |
| **Σ** |  | **2,399369** | **4,057095** |

Тогда .

Статистика Дарбина-Уотсона применяется для проверки нулевой гипотезы  об отсутствии автокорреляции остатков. Альтернативные гипотезы  и  состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам определяются критические значения критерия Дарбина-Уотсона (- нижняя граница признания положительной автокорреляции) и (-верхняя граница признания отсутствия положительной автокорреляции) для заданного числа наблюдений =8, числа независимых переменных модели =1 и уровня значимости =0,05.

=0,76; =1,33.

По этим значениям числовой промежуток  разбивают на пять отрезков.

Положит. Отсутствие Отрицательная

автокорреляция **?** автокорреляции **?** автокорреляция

4 DW

2  

0

0

0

Т.к.  ()  автокорреляция отсутствует с вероятностью =0,95.

(Замечание: Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу .)

**9.** **Вывод:**

Уравнение линейной зависимости расходов на конечное потребление от величины располагаемого дохода приведём вместе с показателями его качества:



Коэффициент детерминации  свидетельствует о высокой доле объяснённой дисперсии (около 96 %) зависимой переменной и о высокой степени приближения исходных данных к построенному уравнению.

Величина *F* – критерия свидетельствует о значимости уравнения в целом, т.е. о наличии значимой статистической связи между переменными.

Значения *t –* статистик для оценок коэффициентов уравнения могут свидетельствовать о значимом отличии от нуля коэффициента регрессии *b* и о возможности исключения из уравнения свободного члена *а*.

Статистика Дарбина-Уотсона позволяет принять гипотезу об отсутствии автокорреляции первого порядка ошибок регрессии и может свидетельствовать об отсутствии зависимости между ними.

**Общим итогом анализа качества построенного уравнения может стать вывод о его хорошем качестве и возможности переоценки на тех же данных с исключением из него свободного члена.**