**Проверка на наличие гетероскедастичности**

**Уравнение парной регрессии**.

**Использование графического метода**.

Этот метод применяют для наглядного изображения формы связи между изучаемыми экономическими показателями. Для этого в прямоугольной системе координат строят график, по оси ординат откладывают индивидуальные значения результативного признака Y, а по оси абсцисс - индивидуальные значения факторного признака X.

Совокупность точек результативного и факторного признаков называется **полем корреляции**.

На основании поля корреляции можно выдвинуть гипотезу (для генеральной совокупности) о том, что связь между всеми возможными значениями X и Y носит линейный характер.

Линейное уравнение регрессии имеет вид y = bx + a

Оценочное уравнение регрессии (построенное по выборочным данным) будет иметь вид y = bx + a + ε, где ei – наблюдаемые значения (оценки) ошибок εi, a и b соответственно оценки параметров α и β регрессионной модели, которые следует найти.

Здесь ε - случайная ошибка (отклонение, возмущение).

Причины существования случайной ошибки:

1. Невключение в регрессионную модель значимых объясняющих переменных;

2. Агрегирование переменных. Например, функция суммарного потребления – это попытка общего выражения совокупности решений отдельных индивидов о расходах. Это лишь аппроксимация отдельных соотношений, которые имеют разные параметры.

3. Неправильное описание структуры модели;

4. Неправильная функциональная спецификация;

5. Ошибки измерения.

Так как отклонения εi для каждого конкретного наблюдения i – случайны и их значения в выборке неизвестны, то:

1) по наблюдениям xi и yi можно получить только оценки параметров α и β

2) Оценками параметров α и β регрессионной модели являются соответственно величины а и b, которые носят случайный характер, т.к. соответствуют случайной выборке;

Для оценки параметров α и β - используют МНК (метод наименьших квадратов).

Метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки относительно случайного члена (ε) и независимой переменной (x).

Формально критерий МНК можно записать так:

S = ∑(yi - y∙i)2 → min

Система нормальных уравнений.

a·n + b·∑x = ∑y

a·∑x + b·∑x2 = ∑y·x

Для расчета параметров регрессии построим расчетную таблицу (табл. 1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | y | x2 | y2 | x∙y |
| 30 | 80 | 900 | 6400 | 2400 |
| 40 | 88 | 1600 | 7744 | 3520 |
| 42 | 85 | 1764 | 7225 | 3570 |
| 43 | 87 | 1849 | 7569 | 3741 |
| 45 | 89 | 2025 | 7921 | 4005 |
| 47 | 91 | 2209 | 8281 | 4277 |
| 49 | 93 | 2401 | 8649 | 4557 |
| 45 | 89 | 2025 | 7921 | 4005 |
| 51 | 95 | 2601 | 9025 | 4845 |
| 52 | 99 | 2704 | 9801 | 5148 |
| 55 | 101 | 3025 | 10201 | 5555 |
| 53 | 102 | 2809 | 10404 | 5406 |
| 56 | 99 | 3136 | 9801 | 5544 |
| 57 | 103 | 3249 | 10609 | 5871 |
| 58 | 105 | 3364 | 11025 | 6090 |
| 55 | 104 | 3025 | 10816 | 5720 |
| 59 | 109 | 3481 | 11881 | 6431 |
| 61 | 110 | 3721 | 12100 | 6710 |
| 62 | 108 | 3844 | 11664 | 6696 |
| 63 | 111 | 3969 | 12321 | 6993 |
| 1023 | 1948 | 53701 | 191358 | 101084 |

Для наших данных система уравнений имеет вид

20a + 1023·b = 1948

1023·a + 53701·b = 101084

Домножим уравнение (1) системы на (-51.15), получим систему, которую решим методом алгебраического сложения.

-1023a -52326.45 b = -99640.2

1023∙a + 53701∙b = 101084

Получаем:

1374.55∙b = 1443.8

Откуда b = 1.0504

Теперь найдем коэффициент «a» из уравнения (1):

20a + 1023∙b = 1948

20a + 1023∙1.0504 = 1948

20a = 873.461

a = 43.6731

Получаем эмпирические коэффициенты регрессии: b = 1.0504, a = 43.6731

Уравнение регрессии (эмпирическое уравнение регрессии):

y = 1.0504 x + 43.6731

Эмпирические коэффициенты регрессии a и b являются лишь оценками теоретических коэффициентов βi, а само уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных.

**1. Параметры уравнения регрессии**.

Выборочные средние.

Выборочные дисперсии:

Среднеквадратическое отклонение

Коэффициент корреляции b можно находить по формуле, не решая систему непосредственно:

**1.1. Коэффициент корреляции**.

*Ковариация*.

Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

Линейный коэффициент корреляции принимает значения от –1 до +1.

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются по шкале Чеддока:

0.1 < rxy < 0.3: слабая;

0.3 < rxy < 0.5: умеренная;

0.5 < rxy < 0.7: заметная;

0.7 < rxy < 0.9: высокая;

0.9 < rxy < 1: весьма высокая;

В нашем примере связь между признаком Y и фактором X весьма высокая и прямая.

Кроме того, коэффициент линейной парной корреляции может быть определен через коэффициент регрессии b:

**1.2. Уравнение регрессии** (оценка уравнения регрессии).

Линейное уравнение регрессии имеет вид y = 1.05∙x + 43.673

Коэффициентам уравнения линейной регрессии можно придать экономический смысл.

Коэффициент регрессии b = 1.05 показывает среднее изменение результативного показателя (в единицах измерения у) с повышением или понижением величины фактора х на единицу его измерения. В данном примере с увеличением на 1 единицу y повышается в среднем на 1.05.

Коэффициент a = 43.673 формально показывает прогнозируемый уровень у, но только в том случае, если х=0 находится близко с выборочными значениями.

Но если х=0 находится далеко от выборочных значений х, то буквальная интерпретация может привести к неверным результатам, и даже если линия регрессии довольно точно описывает значения наблюдаемой выборки, нет гарантий, что также будет при экстраполяции влево или вправо.

Подставив в уравнение регрессии соответствующие значения х, можно определить выровненные (предсказанные) значения результативного показателя y(x) для каждого наблюдения.

Связь между у и х определяет знак коэффициента регрессии b (если > 0 – прямая связь, иначе - обратная). В нашем примере связь прямая.

Для оценки качества параметров регрессии построим расчетную таблицу (табл. 2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | y | y(x) |  |  |
| 30 | 80 | 75.184 | 302.76 | 23.189 |
| 40 | 88 | 85.688 | 88.36 | 5.344 |
| 42 | 85 | 87.789 | 153.76 | 7.779 |
| 43 | 87 | 88.839 | 108.16 | 3.383 |
| 45 | 89 | 90.94 | 70.56 | 3.764 |
| 47 | 91 | 93.041 | 40.96 | 4.165 |
| 49 | 93 | 95.142 | 19.36 | 4.587 |
| 45 | 89 | 90.94 | 70.56 | 3.764 |
| 51 | 95 | 97.242 | 5.76 | 5.029 |
| 52 | 99 | 98.293 | 2.56 | 0.5 |
| 55 | 101 | 101.444 | 12.96 | 0.197 |
| 53 | 102 | 99.343 | 21.16 | 7.059 |
| 56 | 99 | 102.494 | 2.56 | 12.21 |
| 57 | 103 | 103.545 | 31.36 | 0.297 |
| 58 | 105 | 104.595 | 57.76 | 0.164 |
| 55 | 104 | 101.444 | 43.56 | 6.533 |
| 59 | 109 | 105.645 | 134.56 | 11.253 |
| 61 | 110 | 107.746 | 158.76 | 5.079 |
| 62 | 108 | 108.797 | 112.36 | 0.635 |
| 63 | 111 | 109.847 | 184.96 | 1.329 |
| 1023 | 1948 | 1948 | 1622.8 | 106.261 |

**Выводы**.

Изучена зависимость Y от X. На этапе спецификации была выбрана парная линейная регрессия. Оценены её параметры методом наименьших квадратов: y = 1.05∙x + 43.673 Возможна экономическая интерпретация параметров модели - увеличение X на 1 ед.изм. приводит к увеличению Y в среднем на 1.05 ед.изм.

Линейный коэффициент корреляции равен 0.967, следовательно, связь между признаком Y и фактором X весьма высокая и прямая.

**Проверка предпосылок МНК**.

Уравнение регрессии было построено с помощью метода наименьших квадратов (МНК). Данный метод можно применять, если выполняются ряд условий.

**Проверка наличия гетероскедастичности**.

**1) Методом графического анализа остатков**.

В этом случае по оси абсцисс откладываются значения объясняющей переменной X, а по оси ординат либо отклонения ei, либо их квадраты ei2.

Если имеется определенная связь между отклонениями, то гетероскедастичность имеет место. Отсутствие зависимости скорее всего будет свидетельствовать об отсутствии гетероскедастичности.

**2) При помощи теста ранговой корреляции Спирмена**.

**Коэффициент ранговой корреляции Спирмена**.

Присвоим ранги признаку |ei| и фактору X.

|  |  |  |  |
| --- | --- | --- | --- |
| X | |ei| | ранг X, dx | ранг |ei|, dy |
| 30 | 4.816 | 1 | 20 |
| 40 | 2.312 | 2 | 14 |
| 42 | 2.789 | 3 | 17 |
| 43 | 1.839 | 4 | 7 |
| 45 | 1.94 | 5 | 8 |
| 47 | 2.041 | 7 | 10 |
| 49 | 2.142 | 8 | 11 |
| 45 | 1.94 | 5 | 8 |
| 51 | 2.242 | 9 | 12 |
| 52 | 0.707 | 10 | 4 |
| 55 | 0.444 | 12 | 2 |
| 53 | 2.657 | 11 | 16 |
| 56 | 3.494 | 14 | 19 |
| 57 | 0.545 | 15 | 3 |
| 58 | 0.405 | 16 | 1 |
| 55 | 2.556 | 12 | 15 |
| 59 | 3.355 | 17 | 18 |
| 61 | 2.254 | 18 | 13 |
| 62 | 0.797 | 19 | 5 |
| 63 | 1.153 | 20 | 6 |

Так как в матрице имеются связанные ранги (одинаковый ранговый номер) 1-го ряда, произведем их переформирование. Переформирование рангов производиться без изменения важности ранга, то есть между ранговыми номерами должны сохраниться соответствующие соотношения (больше, меньше или равно). Также не рекомендуется ставить ранг выше 1 и ниже значения равного количеству параметров (в данном случае n = 20). Переформирование рангов производится в табл.

|  |  |  |
| --- | --- | --- |
| Номера мест в упорядоченном ряду | Расположение факторов по оценке эксперта | Новые ранги |
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5.5 |
| 6 | 5 | 5.5 |
| 7 | 7 | 7 |
| 8 | 8 | 8 |
| 9 | 9 | 9 |
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12.5 |
| 13 | 12 | 12.5 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |

Так как в матрице имеются связанные ранги 2-го ряда, произведем их переформирование. Переформирование рангов производится в табл.

|  |  |  |
| --- | --- | --- |
| Номера мест в упорядоченном ряду | Расположение факторов по оценке эксперта | Новые ранги |
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 7 | 7 |
| 8 | 8 | 8.5 |
| 9 | 8 | 8.5 |
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12 |
| 13 | 13 | 13 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |

Матрица рангов.

|  |  |  |
| --- | --- | --- |
| ранг X, dx | ранг |ei|, dy | (dx - dy)2 |
| 1 | 20 | 361 |
| 2 | 14 | 144 |
| 3 | 17 | 196 |
| 4 | 7 | 9 |
| 5.5 | 8.5 | 9 |
| 7 | 10 | 9 |
| 8 | 11 | 9 |
| 5.5 | 8.5 | 9 |
| 9 | 12 | 9 |
| 10 | 4 | 36 |
| 12.5 | 2 | 110.25 |
| 11 | 16 | 25 |
| 14 | 19 | 25 |
| 15 | 3 | 144 |
| 16 | 1 | 225 |
| 12.5 | 15 | 6.25 |
| 17 | 18 | 1 |
| 18 | 13 | 25 |
| 19 | 5 | 196 |
| 20 | 6 | 196 |
| 210 | 210 | 1744.5 |

Проверка правильности составления матрицы на основе исчисления контрольной суммы:

Сумма по столбцам матрицы равны между собой и контрольной суммы, значит, матрица составлена правильно.

Поскольку среди значений признаков х и у встречается несколько одинаковых, т.е. образуются связанные ранги, то в таком случае коэффициент Спирмена вычисляется как:

p = 1 -

где

j - номера связок по порядку для признака х;

Аj - число одинаковых рангов в j-й связке по х;

k - номера связок по порядку для признака у;

Вk - число одинаковых рангов в k-й связке по у.

A = [(23-2) + (23-2)]/12 = 1

B = [(23-2)]/12 = 0.5

D = A + B = 1 + 0.5 = 1.5

Связь между признаком |ei| и фактором X слабая и обратная

**Оценка коэффициента ранговой корреляции Спирмена.**

Значимость коэффициента ранговой корреляции Спирмена

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента ранговой корреляции Спирмена при конкурирующей гипотезе Hi. p ≠ 0, надо вычислить критическую точку:

где n - объем выборки; p - выборочный коэффициент ранговой корреляции Спирмена: t(α, к) - критическая точка двусторонней критической области, которую находят по таблице критических точек распределения Стьюдента, по уровню значимости α и числу степеней свободы k = n-2.

Если |p| < Тkp - нет оснований отвергнуть нулевую гипотезу. Ранговая корреляционная связь между качественными признаками не значима. Если |p| > Tkp - нулевую гипотезу отвергают. Между качественными признаками существует значимая ранговая корреляционная связь.

По таблице Стьюдента находим t(α/2, k) = (0.01/2;18) = 3.197

Поскольку Tkp > p, то принимаем гипотезу о равенстве 0 коэффициента ранговой корреляции Спирмена. Другими словами, коэффициент ранговой корреляции статистически - не значим и ранговая корреляционная связь между оценками по двум тестам незначимая.

Проверим гипотезу H0: гетероскедастичность отсутствует.

Поскольку 3.197 > 0.72, то гипотеза об отсутствии гетероскедастичности принимается.

**3. Тест Голдфелда-Квандта**.

В данном случае предполагается, что стандартное отклонение σi = σ(εi) пропорционально значению xi переменной X в этом наблюдении, т.е. σ2i = σ2x2i , i = 1,2,…,n.

Тест Голдфелда-Квандта состоит в следующем:

1. Все n наблюдений упорядочиваются по величине X.

2. Вся упорядоченная выборка после этого разбивается на три подвыборки размерностей k,(n-2k),k.

3. Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) и для третьей подвыборки (k последних наблюдений).

4. Для сравнения соответствующих дисперсий строится соответствующая F-статистика:

F = S3/S1

Построенная F-статистика имеет распределение Фишера с числом степеней свободы v1 = v2 = (n – c - 2m)/2.

5. Если F > Fkp, то гипотеза об отсутствии гетероскедастичности отклоняется.

Этот же тест может использоваться при предположении об обратной пропорциональности между σi и значениями объясняющей переменной. При этом статистика Фишера имеет вид:

F = S1/S3

1. Упорядочим все значения по величине X.

2. Находим размер подвыборки k = (20 - 5)/2 = 8.

где c = 4n/15 = 4∙20/15 = 5

3. Оценим регрессию для первой подвыборки.

**Находим параметры уравнения методом наименьших квадратов**.

Система уравнений МНК:

a0n + a1∑x = ∑y

a0∑x + a1∑x2 = ∑y•x

Для наших данных система уравнений имеет вид:

8a0 + 341a1 = 702

341a0 + 14773a1 = 30075

Из первого уравнения выражаем а0 и подставим во второе уравнение

Получаем a0 = 0.64, a1 = 60.47

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x | y | x2 | y2 | x∙y | y(x) | (y-y(x))2 |
| 30 | 80 | 900 | 6400 | 2400 | 79.669 | 0.109 |
| 40 | 88 | 1600 | 7744 | 3520 | 86.07 | 3.725 |
| 42 | 85 | 1764 | 7225 | 3570 | 87.35 | 5.522 |
| 43 | 87 | 1849 | 7569 | 3741 | 87.99 | 0.98 |
| 45 | 89 | 2025 | 7921 | 4005 | 89.27 | 0.073 |
| 45 | 89 | 2025 | 7921 | 4005 | 89.27 | 0.073 |
| 47 | 91 | 2209 | 8281 | 4277 | 90.55 | 0.202 |
| 49 | 93 | 2401 | 8649 | 4557 | 91.83 | 1.368 |
| 341 | 702 | 14773 | 61710 | 30075 | 702 | 12.054 |

Здесь S1 = 12.05

Оценим регрессию для третьей подвыборки.

**Находим параметры уравнения методом наименьших квадратов**.

Система уравнений МНК:

a0n + a1∑x = ∑y

a0∑x + a1∑x2 = ∑y•x

Для наших данных система уравнений имеет вид:

8a0 + 471a1 = 849

471a0 + 27789a1 = 50055

Из первого уравнения выражаем а0 и подставим во второе уравнение

Получаем a0 = 1.19, a1 = 36

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x | y | x2 | y2 | x∙y | y(x) | (y-y(x))2 |
| 55 | 104 | 3025 | 10816 | 5720 | 101.51 | 6.202 |
| 56 | 99 | 3136 | 9801 | 5544 | 102.701 | 13.695 |
| 57 | 103 | 3249 | 10609 | 5871 | 103.892 | 0.795 |
| 58 | 105 | 3364 | 11025 | 6090 | 105.083 | 0.00686 |
| 59 | 109 | 3481 | 11881 | 6431 | 106.274 | 7.432 |
| 61 | 110 | 3721 | 12100 | 6710 | 108.656 | 1.806 |
| 62 | 108 | 3844 | 11664 | 6696 | 109.847 | 3.412 |
| 63 | 111 | 3969 | 12321 | 6993 | 111.038 | 0.00146 |
| 471 | 849 | 27789 | 90217 | 50055 | 849 | 33.35 |

Здесь S3 = 33.35

Число степеней свободы v1 = v2 = (n – c - 2m)/2 = (20 - 5 - 2∙1)/2 = 6.5

Fkp(6.5,6.5) = 12.2464

Строим F-статистику:

F = 33.35/12.05 = 2.77

Поскольку F < Fkp = 12.2464, то гипотеза об отсутствии гетероскедастичности **принимается**.

Решение было получено и оформлено с помощью сервиса:

[Проверка гетероскедастичности и гомоскедастичности](https://math.semestr.ru/corel/heteroskedasticity.php)

Вместе с этой задачей решают также:

[Уравнение множественной регрессии](https://math.semestr.ru/regress/corel.php)

[Коэффициент корреляции Спирмена](https://math.semestr.ru/corel/spirmen.php)

[Выявление тренда методом аналитического выравнивания](https://math.semestr.ru/trend/analis.php)

[Уравнение нелинейной регрессии](https://math.semestr.ru/corel/noncorel.php)