Министерство Образования Российской Федерации

Алтайский Государственный Университет

## Экономический Факультет

## Заочное отделение

## Кафедра «Информационные системы в экономике»

Контрольная работа по предмету:

«Эконометрика»

Выполнил студент

3 курса 211 группы

Неклюдов А.А.

Барнаул 2003 г.

**Выборочная ковариация**

Выборочная ковариация является мерой взаимосвязи между двумя переменными. Данное понятие может быть продемонстрировано на простом примере. Просматривая табличные данные, помещенные в приложении книги: «Введение в эконометрику», Кристофера Доугерти можно увидеть, что в период между 1963 и 1972 гг. потребительский спрос на бензин в США устойчиво повышался. Эта тенденция прекратилась в 1973 г., а затем последовали нерегулярные колебания спроса с незначительным его падением в целом. В табл. 1.1 приведены данные о потребительском спросе и реальных ценах после нефтяного кризиса. (Реальная цена вычисляется путем деления индекса номинальной цены на бензин, на общий индекс потребительских цен и умножением результата на 100, из таблицы дефляторов цен для личных потребительских расходов(1972 = 100%)). Индексы из таблицы дефляторов основаны на данных 1972 г.; таким образом, индекс реальной цены в таблице 1.1 показывает повышение цены бензина относительно общей инфляции начиная с 1972 г.

Таблица 1.1

|  |
| --- |
| Потребительские расходы на бензин и его реальная цена в США  |
| Год | Расходы(млрд. долл., цены 1972 г.) | Индекс реальных цен (1972=100) |
| 1973 | 26,2 | 103,5 |
| 1974 | 24,8 | 127,0 |
| 1975 | 25,6 | 126,0 |
| 1976 | 26,8 | 124,8 |
| 1977 | 27,7 | 124,7 |
| 1978 | 28,3 | 121,6 |
| 1979 | 27,4 | 179,7 |
| 1980 | 25,1 | 188,8 |
| 1981 | 25,2 | 193,6 |
| 1982 | 25,6 | 173,9 |

Можно видеть некоторую отрицательную связь между потребительским спросом на бензин и его реальной ценой. Показатель выборочной ковариации позволяет выразить данную связь единичным числом. Для его вычисления сначала необходимо найти средние значения цены и спроса на бензин. Обозначив цену через p и спрос – через y, находим средние значения p и y, затем для каждого вычисляем отклонение величин p и y от средних и перемножаем их. Проделаем это для всех годов выборки и возьмем среднюю величину, она и будет выборочной ковариацией (Таблица 1.2).

|  |
| --- |
| Таблица 1.2 |
| Наблюдение |  Ценаp | Спросy |  \_(p-p) |  \_(y-y) |  \_ \_(p-p)(y-y) |
| 1973 | 103,5 | 26,2 | -39,86 | -0,07 | 2,79 |
| 1974 | 127,0 | 24,8 | -16,36 | -1,47 | 24,05 |
| 1975 | 126,0 | 25,6 | -17,36 | -0,67 | 11,63 |
| 1976 | 124,8 | 26,8 | -18,56 | 0,53 | -9,84 |
| 1977 | 124,7 | 27,7 | -18,66 | 1,43 | -26,68 |
| 1978 | 121,6 | 28,3 | -21,76 | 2,03 | -44,17 |
| 1979 | 149,7 | 27,4 | 6,34 | 1,13 | 7,16 |
| 1980 | 188,8 | 25,1 | 45,44 | -1,17 | -53,16 |
| 1981 | 193,6 | 25,2 | 50,24 | -1,07 | -53,76 |
| 1982 | 173,9 | 25,6 | 30,54 | -0,67 | -20,46 |
| Сумма: | 1433,6 | 262,7 |  |  | -162,44 |
| Среднее: | 143,36 | 26,27 |  |  | -16,24 |

Итак, при наличии n наблюдений двух переменных (x и y) выборочная ковариация задается формулой:

Cov(x,y) = 1/n\*(xi-x)(yi-y) = 1/n{(xi-x)(yi-y)+…+(xn-x)(yn-y)}

Следует отметить, что в данном примере ковариация отрицательна. Так это и должно быть. Отрицательная связь в данном примере выражается отрицательной ковариацией, а положительная связь – положительной ковариацией.

Так, например, в наблюдении за 1979 г. (p-pсредн.) = 6,34, (y-yсредн.) = 1,13, а поэтому и их произведение положительно и равно 7,16, в этом наблюдении значения реальной цены и спроса выше соответствующих средних значений следовательно, наблюдение дает положительный вклад в ковариацию.

В наблюдении за 1978 г. реальная цена ниже средней, а спрос выше среднего, поэтому (p-pсредн.) отрицательно, (y-yсредн.) положительно, их произведение отрицательно, и наблюдение вносит отрицательный вклад в ковариацию.

В наблюдении за 1974 г., как реальная цена, так и спрос, ниже своих средних значений, таким образом, (p-pсредн.) и (y-yсредн.) оба являются отрицательными, а их произведение положительно следовательно, наблюдение вносит положительный вклад в ковариацию.

И, наконец, в наблюдении за 1981 г. цена выше средней, а спрос ниже среднего. Таким образом (p-pсредн.) положительно, (y-yсредн.) отрицательно, поэтому (p-pсредн.)(y-yсредн.) отрицательно, и в ковариацию, соответственно, вносится отрицательный вклад.

Несколько основных правил расчета ковариации.

* Правило 1

Если y = v+w, то Cov(x,y) = Cov(x,v)+Cov(x,w).

* Правило 2

Если y = az, где a – константа, то Cov(x,y) = aCov(x,z)

* Правило 3

Если y = a, где a – константа, то Cov(x,y) = 0

##### Демонстрация правила 1

Возьмем данные по шести семьям (домохозяйствам), приведенные в таблице 1.3: общий годовой доход (x); расходы на питание и одежду (y); расходы на питание (v) и расходы на одежду (w). Естественно, y равняется сумме v и w. Указанную в таблице величину z рассмотрим для демонстрации правила 2.

Таблица 1.3

|  |
| --- |
|  |
| Семья | Доход семьи(x) | Расходы на питание и одежду(y) | Расходы на питание(v) | Расходы на одежду(w) | Вторая выборка: расходы семьи на питание и одежду(z) |
| 1 | 3000 | 1100 | 850 | 250 | 2200 |
| 2 | 2500 | 850 | 700 | 150 | 1700 |
| 3 | 4000 | 1200 | 950 | 250 | 2400 |
| 4 | 6000 | 1600 | 1150 | 450 | 3200 |
| 5 | 3300 | 1000 | 800 | 200 | 2000 |
| 6 | 4500 | 1300 | 950 | 350 | 2600 |
| Сумма: | 23300 | 7050 | 5400 | 1650 | 14100 |
| Среднее: | 3883 | 1175 | 900 | 275 | 2350 |

В таблице 1.4 величины (x-x), (y-y), (v-v) и (w-w) вычисляются для каждой семьи. Отсюда получаем (x-xсредн.)(y-yсредн.), (x-xсредн.)(v-vсредн.) и (x-xсредн.)(w-wсредн.) для каждой семьи. Cov(x,y) получается как среднее из величин (x-xсредн.)(y-yсредн.) и равняется 266250. Cov(x,v) равна 157500 и Cov(x,w) = 108750. Следовательно, Cov(x,y) является суммой Cov(x,v) и Cov(x,w).

Таблица 1.4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Семья |  \_x-x |  \_y-y |  \_ \_(x-x)(y-y) |  \_(v-v) |  \_ \_(x-x)(v-v) |  \_(w-w) |  \_ \_(x-x)(w-w) |
| 1 | -883 | -75 | 66250 | -50 | 44167 | -25 | 22083 |
| 2 | -1383 | -325 | 449583 | -200 | 276667 | -125 | 172917 |
| 3 | 117 | 25 | 2917 | 50 | 5833 | -25 | -2917 |
| 4 | 2117 | 425 | 899583 | 200 | 529167 | 175 | 370416 |
| 5 | -583 | -175 | 102083 | -100 | 58333 | -75 | 43750 |
| 6 | 617 | 125 | 77083 | 50 | 30833 | 75 | 46250 |
| Сумма: |  |  | 1597500 |  | 945000 |  | 652500 |
| Среднее: |  |  | 266250 |  | 157500 |  | 108750 |

Демонстрация правила 2

В таблице 1.3 последняя колонка (z) дает расходы на питание и одежду для второго множества из 6 семей. Каждое наблюдение z фактически представляет собой удвоенное значение y. Предполагается, что значения величины x для второго набора семей являются такими же, как и ранее. Для вычисления Cov(x,z) необходимы значения (x-xсредн.), а также (z-zсредн.)

Таблица 1.5

|  |  |  |  |
| --- | --- | --- | --- |
| Семья |  (x-x) |  (z-z) |  (x-x)(z-z) |
| 1 | -883 | -150 | 132500 |
| 2 | -1383 | -650 | 899167 |
| 3 | 117 | 50 | 5833 |
| 4 | 2117 | 850 | 1700167 |
| 5 | -583 | -350 | 204167 |
| 6 | 617 | 250 | 154167 |
| Сумма: |  |  | 3195000 |
| Среднее: |  |  | 532500 |

Из таблицы 1.5 можно видеть, что Cov(x,z) равна 532500, что в точности равно удвоенной Cov(x,y).

##### Демонстрация правила 3

Допустим, что каждая семья в выборке имеет по два взрослых человека, и предположим, что по недоразумению мы решили вычислить ковариацию между общим доходом (x) и числом взрослых в семье (a). Естественно, что a1=a2=…=a6=2. Таким образом, aсредн .= 2. Отсюда для каждой семьи (a-aсредн.) = 0 и, следовательно, (x-xсредн.)(a-aсредн.) = 0. Поэтому Cov(x,a) = 0.

Теоретическая ковариация

Если x и y – случайные величины, теоретическая ковариация xy определяется как математическое ожидание произведения отклонений величин от их средних значений:

pop.cov(x,y) =xy = E{(x)(y-y)}

Если теоретическая ковариация неизвестна, то для ее оценки может использована выборочная ковариация, вычисленная по ряду наблюдений. К сожалению такая оценка ,будет иметь отрицательное смещение.

Если x и y независимы, то их теоретическая ковариация равна нулю, поскольку:

E{(xx)(yy)} = E(xx)(yy) = 0\*0

Выборочная дисперсия.

Для выборки из n наблюдений x1,…,xn выборочная дисперсия определяется как среднеквадратичное отклонение в выбоке:

Var(x) = 1/n(x-x)2

Замечание. Определеннаятаким образом выборочная дисперсия представляет собой смещенную оценку теоретической дисперсии s2, которая определяется как:

1/(n-1)(x-x)2, является несмещенной оценкой 2. Отсюда следует, что ожидаемое значение величины Var(x) равно [(n-1)/n]2 и, следовательно, она имеет отрицательное смещение. Отметим, что если размер выборки n становится большим, то (n-1)/n стремится к единице и, таким образом, математическое ожидание величины Var(x) стремится к 2.

Правила расчета дисперсии.

* Правило 1

Если y = v+w, то Var(y) = Var(v)+Var(w)+2Cov(v,w)

* Правило 2

Если y = az, где a является постоянной, то Var(y) = a2Var(z)

* Правило 3

Если y = a, где a является постоянной, то Var(y) = 0

* Правило 4

Если y = v+a, где a является постоянной, то Var(y) = Var(v)

Следует заметить, что дисперсия переменной x может рассматриваться как ковариация между двумя величинами x:

Var(x) = 1/n\*(xi-x)2 = 1/n\*(xi-x)(xi-x) = Cov(x,x)

Учитывая это равенство, можно воспользоваться правилами расчета выборочной ковариации, чтобы вывести правила расчета дисперсии.

Коэффициент корреляции

Рассматривая ковариацию нельзя не отметить, что она является не особенно хорошим измерителем взаимосвязи между величинами. Более точной мерой зависимости является тесно связанный с ней коэффициент корреляции. Подобно дисперсии и ковариации, коэффициент корреляции имеет две формы – теоретическую и выборочную.

Для переменных x и y теоретический коэффициент корреляции определяется как:

x,y = pop.cov(x,y) / pop.var(x)pop.var(y) = x,y / x2y2 var(y)

Если x и y независимы, то  равно нулю, т.к. равна нулю теоретическая ковариация. Если между переменными существует, то x,y, а следовательно x,y будут положительными. Если существует строгая положительная линейная завистмость, то x,y примет максимальное значение равное 1. Аналогичным образом при отрицательной зависимости x,y будет отрицательным с минимальным значением –1.

Выборочный коэффициент корреляции r равен:

rx,y = (n/(n-1))Cov(x,y) / (n/(n-1))Var(x)(n/(n-1))Var(y)

Множители n/(n-1) сокращаются, поэтому можно определить выборочную корреляцию как:

rx,y = Cov(x,y) / Var(x)Var(y)

Подобно величине , r принимает максимальное значение, равное единице, которая получается при строгой линейной зависимости между выборочными значениями x и y. Аналогичным образом r принимает минимальное значение –1, когда существует линейная отрицательная зависимость. Величина r = 0 показывает, что зависимость между наблюдениями x и y в выборке отсутствует. Однако, тот факт, что r = 0, необязательно означает, что, и наоборот.

Для вычисления выборочного коэффициента корреляции используем пример о спросе на бензин. Данные представлены в таблице 1.1.

Cov(p,y) = – 16,24 (см. табл. 1.2), поэтому теперь необходимо найти значения Var(p) и Var(y) (см. табл. 1.6 на следующей странице). В последних двух колонках таблицы 1.6 можно найти, что Var(p)=888,58 Var(y)=1,33. Следовательно:

r = –16.24 / 888,58 \* 1,33 = – 16,24 / 34,38 = – 0,47

Таблица 1.6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Наблюдение | p | y |  (p-p) |  (y-y)2 |  (p-p)2 |  (y-y)2 |
| 1 | 103,5 | 26,2 | -39,86 | -0,07 | 1588,82 | 0,01 |
| 2 | 127,0 | 24,8 | -16,36 | -1,47 | 267,65 | 2,16 |
| 3 | 126,0 | 25,6 | -17,36 | -0,67 | 301,37 | 0,45 |
| 4 | 124,8 | 26,8 | -18,56 | 0,53 | 344,47 | 0,28 |
| 5 | 124,7 | 27,7 | -18,66 | 1,43 | 348,20 | 2,05 |
| 6 | 121,6 | 28,3 | -21,76 | 2,03 | 473,50 | 4,12 |
| 7 | 149,7 | 27,4 | 6,34 | 1,13 | 40,20 | 1,28 |
| 8 | 188,8 | 25,1 | 45,44 | -1,17 | 2064,79 | 1,37 |
| 9 | 193,6 | 25,2 | 50,24 | -1,07 | 2524,06 | 1,15 |
| 10 | 173,9 | 25,6 | 30,54 | -0,67 | 932,69 | 0,45 |
| Сумма: | 1433,6 | 262,7 |  |  | 8885,75 | 13,30 |
| Среднее: | 143,36 | 26,27 |  |  | 888,58 | 1,33 |

Почему ковариация не является хорошей мерой связи?

Коэффициент корреляции является более подходящим измерителем зависимости, чем ковариация. Основная причина этого заключается в том, что ковариация зависит от единиц, в которых измеряются переменные x и y, в то время как коэффициент корреляции есть величина безразмерная.

Возвращаясь к примеру со спросом на бензин, если при вычислении индекса реальных цен в качестве базового года взять 1980 г. вместо 1972 г., то в этом случае ковариация изменится, а коэффициент корреляции – нет.

При использовании 1972 г. вкачестве базового года индекс реальных цен для 1980 г. составил 188,8. Если теперь принять этот индекс за 100 для 1980 г., то нужно пересчитать ряды путем умножения на коэффициент 100/188,8 = 0,53. Новые ряды представлены во второй колонке таблицы 1.7 и будут обозначены через P. Величина P численно меньше, чем p.

Так как отдельное наблюдение ряда цен было пересчитано с коэффициентом 0,53 то отсюда следует, что и среднее значение за выборочный период (Pсредн.) пересчитывается с этим коэффициентом. Следовательно, в году t:

Pt – P = 0,53pt – 0,53p = 0,53(pt – p)

Это означает, что в году t:

(P – P)(y – y) = 0,53(p – p)(y – y),

и, следовательно, Cov(P,y) = 0,53Cov(p,y). Однако на коэффициент корреляции это изменение не повлияет. Коэффициент корреляции для P и y будет равен:

rp,y = Cov(P,y) / Var(P)Var(y)

Таблица 1.7

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Наблюдение | P | y | P-P | y-y | (P-P)2 | (y-y)2 | (P-P)(y-y) |
| 1973 | 54,82 | 26,2 | -21,11 | -0,07 | 445,73 | 0,01 | 1,48 |
| 1974 | 67,27 | 24,8 | -8,67 | -1,47 | 75,09 | 2,16 | 12,74 |
| 1975 | 66,74 | 25,6 | -9,20 | -0,67 | 84,55 | 0,45 | 6,16 |
| 1976 | 66,10 | 26,8 | -9,38 | 0,53 | 96,64 | 0,28 | -5,21 |
| 1977 | 66,05 | 27,7 | -9,88 | 1,43 | 97,68 | 2,05 | -14,13 |
| 1978 | 64,41 | 28,3 | -11,53 | 2,03 | 132,84 | 4,12 | -23,40 |
| 1979 | 79,29 | 27,4 | 3,36 | 1,13 | 11,28 | 1,28 | 3,80 |
| 1980 | 100,00 | 25,1 | 24,07 | -1,17 | 579,26 | 1,37 | -28,16 |
| 1981 | 102,54 | 25,2 | 26,61 | -1,07 | 708,10 | 1,15 | -28,47 |
| 1982 | 92,11 | 25,6 | 16,18 | -0,67 | 261,66 | 0,45 | -10,84 |
| Сумма: | 759,32 | 262,7 |  |  | 2492,28 | 13,30 | -86,04 |
| Среднее: | 75,93 | 26,27 |  |  | 249,23 | 1,33 | -8,60 |