

## ОЦЕНИВАНИЕ ВЕРОЯТНОСТИ СЛУЧАЙНОГО СОВПАДЕНИЯ РОЗНИЧНЫХ ЦЕН

Афанасьев Д.А.,

Научный руководитель Крупкина Т.В.  
Сибирский федеральный университет

У ряда индивидуальных предпринимателей и ООО цена на различные виды топлива за некоторый период времени были совпадающими или очень близкими, изменение цен в названный период осуществлялось ими синхронно и единообразно. Вместе с тем ООО и ИП закупали нефтепродукты у различных поставщиков и по различным ценам. Необходимо было оценить вероятность случайного совпадения розничных цен с учетом объективных причин повышения и понижения цен, а также с учетом закупочных цен у каждого участника рынка.

Для исследования статистической природы были рассмотрены распределения величин (*слайд 3*), где  $k_{ci}$  — закупочная цена  $i$ -го предприятия на единицу  $j$ -го продукта,  $\varepsilon_{ta}$  — розничная цена  $i$ -го предприятия на единицу  $j$ -го продукта,  $t$  — неслучайный временной параметр,  $T$  — рассматриваемый промежуток времени.

Во всем массиве исходных данных закупочные цены принимали множество различных значений, зачастую отличающиеся буквально на 1-2 копейки, в то время, как розничные цены принимали всего лишь несколько значений. Из этого можно сделать вывод, что наша статистическая модель  $\langle F(k_{ci}), F(\varepsilon_{ta}) \rangle$  имеет вид  $\langle F(\text{непрерывное}), F(\text{дискретное}) \rangle$

Был проведен корреляционный анализ для каждой пары  $(i, j)$ , рассчитаны коэффициенты корреляции между закупочными и розничными ценами одного и того же предприятия и коэффициент корреляции между розничными ценами различных предприятий, а также коэффициент корреляции между закупочными ценами различных предприятий. (*слайд 4*) В таблице мы видим значение коэффициента корреляции между закупочными ценами ( $k_{ci}$ ) и розничными ценами ( $\varepsilon_{ta}$ ) по  $j$ -му продукту для предприятий с номерами  $i$  и  $k$ . Что характерно, розничные цены у всех участников рынка были коррелированы сильнее, чем закупочные и розничные цены у любого участника.

Безусловно, имеют место общие закономерности изменения розничных и закупочных цен. Так как все розничные цены связаны общим моментом времени  $t$ , то вполне возможно, что для синхронного изменения цен имелись и объективные внешние причины. Тем не менее, графики закупочных цен различных предприятий схожи гораздо меньше, чем графики их розничных цен. (*слайд 5*).

Рассмотрим  $j$ -ый продукт, продаваемый  $m$  субъектами в течении  $N$  дней. Каждый день наблюдений может быть рассмотрен как испытание, где успехом является совпадение цен у рассматриваемых предприятий. (*слайд 6*) В предположении независимости испытания образуют схему Бернулли, и, следовательно, число успехов имеет биномиальное распределение  $B(N, p)$ , где  $N$  — число наблюдений,  $p$  — вероятность успеха в одном испытании. Если вероятность  $P(m \geq m_0) = \text{сумма} \dots$  окажется меньше уровня значимости, то гипотезу о независимости следует отклонить и признать совпадение цен не случайным. Надо оценить параметр  $p$ . Если вероятности совпадения в различные дни различаются, то имеет место закон больших чисел в форме Пуассона:  $m/n - \text{сумма}(p(k))/n \rightarrow 0$  по вероятности, где  $m$  — число успехов,  $m/n$  — частота. (*слайд 7*) Поэтому в качестве параметра схемы Бернулли берем  $p = \text{сумма}(p(k))/N$ , где  $p(k)$  — вероятность совпадения цен у всех предприятий в  $k$ -ый день. Но у каждого предприятия действовала индивидуальная закупочная цена в каждый день, следовательно (последняя формула на слайде).

Рассмотрим теперь конкретное предприятие, продающее  $j$ -ый продукт. Для статистического анализа нам потребуется группировка. Сгруппируем закупочные цены предприятий с некоторым интервалом  $h$ . Были испробованы несколько способов

группировки(слайд 8). Данные сгруппированы за 137 дней, рассматривался период в 151 день, но в 14 из них были удалены из-за пропусков в данных по какому-либо предприятию

Для каждого интервала была составлена таблица розничных цен(слайд 9) где частота определялась как отношения числа дней, в которые была зафиксирована текущая розничная цена, к общему числу дней данного интервала закупочной цены.

Таким образом, в предположении независимости цен предприятий оценка вероятности совпадения цен равна произведению частот, т. е. применяется теорема умножения вероятностей. Расчеты проводятся отдельно для каждого дня совпадений по всей совокупности или по репрезентативной выборке. Условные частоты берутся из таблиц с учетом закупочных цен каждого субъекта в рассматриваемый день. Средняя вероятность совпадения, равная 0,434, принимается за оценку вероятности совпадения цен у субъектов А, Б, С в один день. По каждому продаваемому продукту оценивание параметров проводится отдельно.

Для проведения расчетов по модели разработано программное обеспечение в кросс-платформенном инструментарии QT на языке программирования C++.(слайд 10)

На основе проделанных расчетов была найдена вероятность того, что из определенного количества дней будет наблюдаемое число совпадений. Например, из рассматриваемых 137 дней совпадения цен у трех субъектов наблюдались в 88 дней. Это примерно в полтора раза превышает среднее ожидаемое количество совпадений при условии независимости ( $137 \cdot 0,434 \approx 59$ ). Рассчитаем вероятность не менее 88 совпадений. Число совпадений имеет биномиальное распределение  $B(137; 0,434)$ . Расчет дал  $P(m \geq 88) \approx 0,00000030$ .

Используемая оценка  $p=0,434$  была найдена по данным, приведенным в таблице ранее, т. е. по выборке из 10 дней (Слайд 11). При расчетов по всей совокупности данных была получена оценка 0,499. В этом случае вероятность случайного совпадения розничных цен в 88 случаях из 137 будет равна 0,00039624. В этом случае гипотезу о независимости следует отвергнуть и признать совпадение цен зависимым.

Устойчивость метода достигается путем обработки одних и тех же данных несколько раз с разной величиной шага. Я посчитал необходимым опробовать метод при размере шагов 0.5, 1 и 2. По результатам обработки данных стало видно, что если все три полученные значения оказываются далеко за пределами уровня значимости, то можно быть уверенным в зависимости испытаний. В противном же случае, если величины независимы, то, как минимум, 2 из 3 не превышают уровня значимости и величины признаются независимы. Многократное использование данной методики показывает её устойчивость.

Аналогичные расчеты были проведены по другим продуктам и получены схожие результаты. Кроме того, при расчете возник один интересный случай, так как у одного из предприятий уровень цены постоянно устанавливался на 1 рубль ниже совпадающей цены других предприятий. Данная методика была неприменима в таком случае, однако коэффициент корреляции между ценами в этом случае составляет 1, что указывает на однозначную линейную связь.