“Процедура Экспресс-оценки”

Канонические определения рыночной стоимости объекта оценки трактуют ее как *наиболее вероятную величину цены, по которой данный объект оценки может быть отчужден на открытом рынке в условиях конкуренции*, указывая тем самым на статистическую ее природу [1, с. 2].

С точки зрения математической статистки стоимость, как *случайная величина (СВ)*, рассчитывается на основе значений цен объектов-аналогов x_i, $i = 1, \ldots, n$, понимаемых как n ее независимых наблюдений. В качестве генеральной совокупности выступают цены всех объектов на рассматриваемом сегменте рынка, а стоимость объекта оценки получают в результате обработки доступной оценщику выборки значений из генеральной совокупности [1, с. 2].

В теории и практике оценки обычно в качестве показателя рыночной стоимости используют ее *математическое ожидание*, оценку которого получают расчетом выборочного среднего, сопровождая его, как правило, оценкой точности в виде границ доверительного интервала [1, с. 2].

Достоверность этой оценки должна быть подтверждена [1, с. 3]:

1) доказательством отсутствия в выборке аномальных, резко выделяющихся данных, наличие которых может существенно исказить статистические оценки результирующего значения стоимости;

2) проверкой гипотезы о нормальном (Гауссовском) распределении значений (или, по крайней мере, гипотезы о симметричности [1, с. 8]).

Нормальный закон распределения СВ [5, с. 76 - 79]

Аналитически ***нормальное распределение*** описывается функцией плотности вероятности вида

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} .$$

Предполагается, что переменная x может принимать бесконечно большие и бесконечно малые значения, количество измерений бесконечно, а интервал квантования пренебрежимо мал.

По этой формуле при различных значениях математического ожидания (a) и стандартного отклонения (σ) получается семейство нормальных кривых.

Стандартное или ***единичное нормальное распределение*** с $a = 0$ и $\sigma = 1$ (обозначение $N(z;0;1)$):

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} .$$

Стандартное нормальное распределение удобно тем, что любое другое нормальное распределение может быть сведено к нему путем процедуры стандартизации, т.е. заменой каждого измеренного значения соответствующей ***стандартизированной z-величиной***: $z = \frac{x_i - a}{\sigma}$.

Это позволяет анализировать любые нормальные распределения на основе знания характеристик единичного нормального распределения.

Свойством нормального распределения является наличие определенного количества значений СВ, приходящегося на интервалы между значениями σ, обычно это количество измеряется в % от общего числа наблюдений.

Нормальное распределение значений признака наблюдается в тех случаях, когда на величину признака действует множество случайных независимых или слабо зависимых между собой факторов, каждый из которых играет в общей сумме примерно одинаковую и малую роль (т.е. отсутствуют доминирующие факторы). В результате получается, что чаще наблюдаются некоторые средние значения измеряемого параметра, реже крайние, и чем сильнее отличается какое-то значение от среднего, тем реже оно встречается.

Характеристики центральной тенденции распределения значений СВ

(Выборочное) ***Среднее*** ($\overline{X}$) рассчитывается по формулам (в зависимости от способа представления данных):

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{j=1}^{r} n_j m_j = \sum_{j=1}^{r} h_j m_j ,$$

где $n-$ объем выборки;

x_i – наблюдаемые значения $(i = \overline{1, n})$;

r – количество интервалов (групп) в сгруппированном ряду данных;

m_j – середины интервалов $(j = \overline{1, r})$;

n_j – частоты; $h_j = n_j / n$ – относительные частоты.

Медиана (Md) (середина) – это такое значение СВ, которое делит распределение пополам, т.е. слева и справа от медианного значения находится по 50% наблюдений. В упорядоченном вариационном ряду (выборке) медиана равна:

$$Md = \begin{cases} x_k \left(k = \dfrac{n+1}{2}\right), \textit{ если } n = 2k-1\ (k \in N) - \textit{нечетное}; \\ 0{,}5(x_k + x_{k+1}) \left(k = \dfrac{n}{2}\right), \textit{ если } n = 2k\ (k \in N) - \textit{четное} \end{cases}.$$

Характеристики рассеяния (вариации) значений СВ

Размах вариации (R) – разность между наибольшим ($x_{\max}$) и наименьшим ($x_{\min}$) значениями СВ: $R = x_{\max} - x_{\min}$.

Размах вариации зависит от величины только крайних значений признака. Более точно рассеяние характеризуют показатели, основанные на учете вариации всех значений признака, – среднее абсолютное (линейное) отклонение и стандартное (среднеквадратическое) отклонение.

Среднее абсолютное (линейное) отклонение:

$$d = \frac{\sum_{i=1}^{n} \left| x_i - \overline{X} \right|}{n},$$

где $n-$ объем выборки;

x_i – наблюдаемые значения признака $(i = \overline{1, n})$;

$\overline{X}$ – (выборочное) среднее.

Дисперсия показывает *разброс* значений СВ относительно своего среднего значения, т.е. насколько плотно значения СВ группируются вокруг $\overline{X}$. Чем больше разброс, тем сильнее варьируются результаты наблюдений в данной группе, тем больше индивидуальные различия между значениями признака:

$$D = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n},$$

где $n-$ объем выборки;

x_i – наблюдаемые значения признака $(i = \overline{1, n})$;

$\overline{X}$ – (выборочное) среднее;

Дисперсия D, рассчитанная по приведенной выше формуле , является *смещенной* оценкой дисперсии генерального распределения с отрицательным смещением, равным (D / n). Поэтому для оценки дисперсии по выборочным данным используется ***несмещенная оценка дисперсии (S^2):***

$$S^2 = D - \frac{D}{n} = \frac{n-1}{n} D = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}.$$

(Выборочное) Стандартное (или среднеквадратичное) отклонение (***S***) характеризует *разброс* значений СВ более наглядно, чем дисперсия, т.к. его размерность соответствует размерности измеряемой величины, в отличие от “квадратного” размера дисперсии:

$$S = \sqrt{S^2} = \sqrt{\frac{n-1}{n} D} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}.$$

Коэффициент вариации позволяет сравнивать вариативность случайных величин в разных совокупностях и/или разной природы, т.к. вообще не имеет размерности:

$$V = \frac{S}{\overline{X}} (\cdot 100\%).$$

По величине *коэффициента вариации* можно также судить об интенсивности рассеяния значений признака, а, следовательно, и ***об однородности*** состава изучаемой совокупности. Чем больше величина коэффициента вариации, тем больше разброс значений признака вокруг средней, тем больше неоднородность совокупности.

Интервальное оценивание и стандартные ошибки

Стандартная ошибка означает величину стандартного отклонения распределения данной статистики относительно ее среднего значения при условии извлечения бесконечного числа выборок некоторого объема.

Доверительным интервалом называется интервал, который покрывает неизвестный параметр распределения с заданной (*доверительной*) *вероятностью.*

Доверительной вероятностью (доверительным коэффициентом) неизвестного параметра называется вероятность того, что случайно выбранный интервал из совокупности возможных доверительных интервалов будет содержать этот параметр.

Очень часто при анализе выборочных данных, значения, выходящие за пределы 95%-го (реже 90%-го) доверительного интервала, исключаются из дальнейших расчетов.

Стандартная ошибка среднего $\mu = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$.

95%-ая доверительная вероятность равна $P = 1 - \alpha = 1 - 0.05 = 0.95$, где $\alpha = 0.05 -$ вероятность ошибки.

Нижняя и Верхняя границы 95%-го доверительного интервала для среднего соответственно равны

$$\overline{X} \mp z_{\alpha/2}\, \mu = \overline{X} \mp 1{,}96\, \mu,$$

где $\overline{X}$ – (выборочное) среднее;

μ – стандартная ошибка среднего;

$z_{\alpha/2} = z_{0,05/2} = z_{0,025} = 1{,}96 -$ верхняя процентная точка стандартного нормального распределения [5, с. 496].

Асимметрия и эксцесс

Асимметрия (***As***) характеризует степень несимметричности (скоса) распределения относительно его среднего.

Несмещенная (выборочная) оценка асимметрии рассчитывается по формуле:

$$As = As_Г \frac{\sqrt{n(n-1)}}{n-2} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{X}}{S} \right)^3, \text{ где } As_Г = \frac{\sqrt{n} \sum_{i=1}^{n} \left(x_i - \overline{X}\right)^3}{\sqrt{\left(\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2 \right)^3}}.$$

Для симметричных распределений $As = 0$.

Среднеквадратическое отклонение асимметрии (по выборке) равно

$$S(As) = \sqrt{\frac{6n(n-1)}{(n+1)(n-2)(n+3)}}.$$

Эксцесс (***Ex***) характеризует относительную остроконечность или сглаженность (крутизну) распределения по сравнению с нормальным распределением.

Несмещенная (выборочная) оценка эксцесса рассчитывается по формуле:

$$Ex = \frac{n-1}{(n-2)(n-3)}\left((n+1)Ex_{Г}+6\right) = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\overline{X}}{S}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)},$$

где $Ex_{Г} = \dfrac{n\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^4}{\left(\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2\right)^2}$.

Среднеквадратическое отклонение эксцесса (по выборке) равно

$$S(Ex) = \sqrt{\frac{24n(n-1)^2}{(n-2)(n-3)(n+3)(n+5)}}.$$

Проверка на наличие/отсутствие аномальных значений в выборке

Критерий Граббса (проверка на один выброс) [4, п. 7.3.4.1.]

Для проверки гипотезы об аномальности крайнего (наибольшего или наименьшего) наблюдаемого значения (x_k) рассчитывается эмпирическое значение статистики Граббса:

$$G_{эмп} = \frac{|x_k-\overline{X}|}{S},$$

где $\overline{X}$ – (выборочное) среднее;
S – (выборочное) стандартное отклонение.

По таблице критических значений для критерия Граббса [4, Табл. 5] определяется $G_{кр}(\alpha; n)$.

Если $G_{эмп} > G_{кр}(\alpha; n)$, то с доверительной вероятностью (степенью уверенности) $p = 1-\alpha$ значение x_k признается аномальным, и удаляется из дальнейшего рассмотрения.

Критерий Граббса применяется до тех пор, пока для некоторого значения x_k (наиболее удаленного от среднего) не будет выполнено условие: $G_{эмп} \le G_{кр}(\alpha = 0{,}05; n)$.

Критерий λ* [2, с. 761]

Для проверки гипотезы об аномальности крайнего (наибольшего или наименьшего) наблюдаемого значения (x_k) рассчитывается эмпирическое значение статистики λ*:

$$\lambda^* = \frac{|x_k-\overline{X}|}{S},$$

где $\overline{X}$ – (выборочное) среднее;
S – (выборочное) стандартное отклонение.

По таблице критических значений для статистики λ* [2, с. 761] определяется $\lambda^*_{кр}\ (\alpha = 0{,}05)$.

Если $\lambda^* > \lambda^*_{кр}\ (\alpha = 0{,}05; n)$, то значение x_k признается аномальным, и удаляется из дальнейшего рассмотрения.

Критерий λ* применяется до тех пор, пока для некоторого значения x_k (наиболее удаленного от среднего) не будет выполнено условие: $\lambda^* \le \lambda^*_{кр}\ (\alpha = 0{,}05; n)$.

Проверка гипотезы о нормальном (Гауссовском) распределении значений

1. Проверка по коэффициенту вариации (п. 2.1 [1, с. 7])**:**

Если $V = \frac{S}{\overline{X}} \cdot 100\% > 33\%$, то гипотеза о нормальности распределения данных выборки не подтверждается. Дальнейшую проверку при этом не проводят, так как такие распределения должны преобразовываться с целью уменьшения коэффициента вариации. При обратном соотношении проверку продолжают по оставшимся критериям.

2. Проверка по критерию среднего абсолютного отклонения для выборок объемом $n \leq 120$ (п. 2.1 [1, с. 7])**:**

Для выборки, имеющей приближенно нормальный закон распределения, должно выполняться соотношение

$$\left| \frac{d}{S} - \sqrt{\frac{2}{\pi}} \right| < \frac{0{,}4}{\sqrt{n}}.$$

3. Проверка с помощью показателей асимметрии и эксцесса для выборок объемом $n \geq 20$ (п. 2.4 [1, с. 8]; п. 5.5. [3, с. 61 – 64])**:**

Для выборки, имеющей приближенно нормальный закон распределения, должны выполняться соотношения $\frac{|As|}{S(As)} \leq 3$ **и** $\frac{|Ex|}{S(Ex)} \leq 5$.

Проверка гипотезы о симметричности распределении значений (с помощью показателя асимметрии) (п. 2.4 [п. 1, с. 8])

Выполнение условия симметричности распределения особенно важно, так как в этом случае, даже при отсутствии нормальности можно уверенно использовать среднее $\overline{X}$ в качестве оценки наиболее вероятного значения случайной величины, требуемого определениями рыночной стоимости.

Для симметричного распределения должно выполняться соотношение $\frac{|As|}{S(As)} \leq 3$.

Использованные источники:

1. Анисимова, И.Н. О повышении достоверности оценки рыночной стоимости методом сравнительного анализа [Текст] /И.Н. Анисимова, Н.П. Баринов, С.В. Грибовский. – //Вопросы оценки" – 2002 – № 1 – С. 2 – 10.

2. Владимирский, Б.М. Математика. Общий курс: Учебник [Текст] / Б.М. Владимирский, А.Б. Горстко, Я.М. Ерусалимский. – СПб.: Издательство «Лань», 2004. – 960 с.

3. Статистика: учебник [Текст] / И.И. Елисеева и др.; под редакцией И.И. Елисеевой. – М.: ТК Велби, Изд-во Проспект, 2007. – 448 с.

4. Точность (правильность и прецизионность) методов и результатов измерений. Часть 2. Основной метод определения повторяемости и воспроизводимости стандартного метода измерений (ГОСТ Р ИСО 5725-2-2002) – принят и введен в действие Постановлением Госстандарта России от 23 апреля 2002 г. № 161-ст.

5. Тюрин Ю.Н. Анализ данных на компьютере [Текст] / Ю.Н. Тюрин, А.А. Макаров; под редакцией В.Э. Фигурнова. – М.: ИНФРА-М, 2003. – 544 с.

Математическое обеспечение – Невзорова Татьяна Анатольевна,
Программное обеспечение – к.э.н. Наседкин Сергей Юрьевич.