Методы экспертных оценок. Непосредственная оценка и парное сравнение. Модель Терстоуна. Оценка согласованности для парных сравнений

Методы экспертных оценок являются частью обширной области теории принятия решений, а само **экспертное оценивание** — процедура получения оценки проблемы на основе мнения специалистов (экспертов) с целью последующего принятия решения (выбора).

В случаях чрезвычайной сложности проблемы, ее новизны, недостаточности имеющейся информации, невозможности математической формализации процесса решения приходится обращаться к рекомендациям компетентных специалистов, прекрасно знающих проблему, — к экспертам. Их решение задачи, аргументация, формирование количественных оценок, обработка последних формальными методами получили название метода экспертных оценок.

Существует **две группы экспертных оценок:**

1. Индивидуальные оценки основаны на использовании мнения отдельных экспертов, независимых друг от друга.
2. Коллективные оценки основаны на использовании коллективного мнения экспертов.

Грубо говоря, к первой группе относится оценка статей на хабре, голосование в опросах и т.д., когда каждый эксперт принимает решение самостоятельно. Подбор (отсев) экспертов осуществляется посредством кармы. Именно первая группа превалирует в интернете 2 за счет возможности охвата большего числа экспертов.

Способы измерения объектов

1. **Ранжирование** – это расположение объектов в порядке возрастания или убывания какого-либо присущего им свойства. Ранжирование позволяет выбрать из исследуемой совокупности факторов наиболее существенный.
2. **Парное сравнение** — это установление предпочтения объектов при сравнении всех возможных пар. Здесь не нужно, как при ранжировании, упорядочивать все объекты, необходимо в каждой из пар выявить более значимый объект или установить их равенство.
3. **Непосредственная оценка**. Часто бывает желательным не только упорядочить (ранжировать объекты анализа), но и определить, на сколько один фактор более значим, чем другие. В этом случае диапазон изменения характеристик объекта разбивается на отдельные интервалы, каждому из которых приписывается определенная оценка (балл),

например, от 0 до 10. Именно поэтому метод непосредственной оценки иногда именуют также балльным методом.

В парном сравнении не нужно, как при ранжировании, упорядочивать все объекты, необходимо в каждой из пар выявить более значимый объект или установить их равенство. Парное сравнение можно проводить при большом числе объектов, а также в тех случаях, когда различие между объектами столь незначительно, что практически невыполнимо их ранжирование. При использовании метода чаще всего составляется матрица размером $n_x n$, где n – количество сравниваемых объектов.

	1	2	…	j	…	n
1	a_{11}	a_{12}	…	a_{1j}	…	a_{1n}
2	a_{21}	a_{22}	…	a_{2j}	…	a_{2n}
…	…	…	…	…	…	…
i	a_{i1}	a_{i2}	…	a_{ij}	…	a_{in}
…	…	…	…	…	…	…
n	1_{ni}	a_{n2}	…	a_{nj}	…	a_{nn}

При сравнении объектов матрица заполняется элементами a_{ij} следующим образом (может быть предложена и иная схема заполнения):

- 2, если объект i предпочтительнее объекта j (i > j),
- 1, если установлено равенство объектов (i = j),
- 0, если объект j предпочтительнее объекта i (i < j).
  - В настоящее время во многих методах проведения экспертных оценок предлагается в качестве показателя компетентности эксперта коэффициент:
  - $K_к = \dfrac{K_{зн} + K_а}{2}$, (2.6)
  - где $K_к$ - коэффициент компетентности эксперта;
  - $K_{зн}$ - коэффициент степени знакомства эксперта с обсуждаемой проблемой;
  - $K_а$ - коэффициент аргументированности.
  - Коэффициент степени знакомства с направлением исследований определяется путем самооценки эксперта по десятибалльной шкале. Значения баллов для самооценки следующие:
  - 0 - эксперт не знаком с вопросом;

- 1,2,3 - эксперт плохо знаком с вопросом, но вопрос входит в сферу его интересов;

- 4,5,6 - эксперт удовлетворительно знаком с вопросом, не принимает непосредственного участия в практическом решении вопроса;

- 7,8,9 – эксперт хорошо знаком с вопросом, участвует в практическом решении вопроса;

- 10 – вопрос входит в круг узкой специализации эксперта.

- Эксперту предлагается самому оценить степень своего знакомства с вопросом и подчеркнуть соответствующий балл. Затем этот балл умножается на 0,1, и получаем коэффициент .

- Коэффициент аргументированности учитывает структуру аргументов, послуживших эксперту основанием для определенной оценки. Коэффициент аргументированности предлагается определить в соответствии с таблицей 2.2 путем суммирования значений, отмеченных экспертом в клетках этой таблицы.

- Определив коэффициент компетентности, умножают на него значение оценок экспертов.

- Таблица 2.2 Значения коэффициента аргументированности

-

Источники аргументации	Степень влияния источника аргументации на Ваше мнение		
	высокая	средняя	низкая
Проведенный Вами теоретический анализ	0,3	0,2	0,1
Ваш производственный опыт	0,5	0,4	0.2
Обобщение работ отечественных авторов	0,05	0,05	0,05
Обобщение работ зарубежных авторов	0,05	0,05	0,05
Ваше личное знакомство с состоянием дел за рубежом	0,05	0,05	0,05
Ваша интуиция	0,05	0,05	0,05

<u>Непосредственная оценка</u>. Часто бывает желательным не только упорядочить (ранжировать объекты анализа), но и определить, на сколько один фактор более значим, чем другие. В этом случае диапазон изменения характеристик объекта разбивается на отдельные интервалы, каждому из которых приписывается определенная оценка (балл), *например, от 0 до 10*. Именно поэтому метод непосредственной оценки иногда именуют также <u>балльным методом</u>.

Анализ результатов экспертных оценок

Для анализа результатов применяются различные методы математической статистики. Причем, они могут комбинироваться и варьироваться в зависимости от типа задачи и необходимого результата.

Формирование обобщенной оценки

Итак, пусть группа экспертов оценила какой-либо объект, тогда x_j – оценка j-го эксперта, где m – число экспертов.
Для формирования обобщенной оценки группы экспертов чаще всего используются средние величины. Например, медиана, за которую принимается такая оценка, по отношению к которой число больших оценок равняется числу меньших.

Установление степени согласованности мнений экспертов

В случае участия в опросе нескольких экспертов расхождения в их оценках неизбежны, однако величина этого расхождения имеет важное значение. Групповая оценка может считаться достаточно надежной только при условии хорошей согласованности ответов отдельных специалистов.
Для анализа разброса и согласованности оценок применяются статистические характеристики – меры разброса или статистическая вариация.
Итак, **способы вычисления меры разбрса**:
Вариационный размах

$$R = x_{\max} - x_{\min};$$

Среднее линейное отклонение

$$a = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|,$$

Среднеквадратическое отклонение

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Дисперсия

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2;$$

Коэффициента ранговой корреляции Спирмэна

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_{ij} - x_{ik})^2}{n(n^2-1)} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

Коэффициент (величина ρ) может изменяться в диапазоне от –1 до +1. При полном совпадении оценок коэффициент равен единице. Равенство коэффициента минус единице наблюдается при наибольшем расхождении в мнениях экспертов.
x_{ij} – ранг (**важность**), присвоенный i-му объекту j-ым экспертом, x_{ik} – ранг, присвоенный i-му объекту k-ым экспертом, d_i – разница между рангами, присвоенными i-му объекту.

Коэффициент конкордации Кенделла
Коэффициент может принимать значения в пределах от 0 до 1. При полной согласованности мнений экспертов коэффициент конкордации равен единице при полном разногласии – нулю.

Наиболее реальным является случай частичной согласованности мнений экспертов

Говоря о согласованности мнений экспертов, стоит упомянуть, что ранжирование не подразумевает (или не всегда подразумевает) расстояние. То есть у одного эксперта A>B>C означает, что A>>B>C, а у другого A>B>>C. И всякие корреляции и расчеты средних оценок тут не помогут. Как вариант, считать индекс согласованности. Что-то типо количества противоречивых замкнутых цепочек мнений экспертов (Первый считает, что А лучше Б, второй, что Б лучше С, а третий, что С лучше А) к количеству всех подобных цепочек.

Самые распространенные в настоящее время методы шкалирования субъективных характеристик стимулов, не имеющих прямых физических коррелятов, основаны на модели шкалирования *Терстоуна* (Терстоун, 1927). Но первый шаг в этом направлении сделали *Фуллертон и Кэтелл* (1892), которые предложили подход, преобразующий постулат Фехнера о равенстве "едва заметных различий" в понятие равенства на континууме "равно часто замечаемых различий". Этот подход позволил перейти к оценке стимула, безотносительно к прямому физическому корреляту, но сразу же обнажилась проблема: если один стимул предпочитается второму с частотой *A*, а второй стимул предпочитается третьему с частотой в *1.2A*, то насколько субъективное расстояние между вторым и третьим стимулами больше субъективного расстояния между первым и вторым стимулами?

Торндайк (1910) предлагает решение этой проблемы (и это можно считать вторым шагом к цели), предположив, что разница в субъективных расстояниях пропорциональна различию в единицах стандартного отклонения нормальной кривой, соответствующих двум частотам.

Полное развитие этих идей и представляет собой модель шкалирования Терстоуна. Суть ее заключается в следующем:

1. Данное множество объектов можно *упорядочить в континуум* по какому-либо из параметров, который может служить стимулом, причем этот параметр не обязательно имеет физическую меру. Обозначим ряд стимулов как $1 \dots i \dots n$.

2. Каждый стимул теоретически вызывает у субъекта только один, свой *процесс различения* (обозначим его буквой S). Процессы различения составляют *психологический континуум, или континуум различения* ($D_1 \dots D_i \dots D_n$). Однако вследствие мгновенных флуктуаций организма, данный стимул может

вызвать не только свой процесс различения, но и какие-то соседние. Поэтому, если один и тот же стимул предъявлять много раз, то на психологическом континууме ему будет соответствовать некоторое *распределение процессов различения*. При этом предполагается, что *форма распределения нормальна*.

3. В качестве значения i-го стимула на психологической шкале принимается *среднее* (S_i) распределения процессов различения, а дисперсия распределения рассматривается как *дисперсия различения* (s_i).

4. Предъявление одновременно пары стимулов вызывает два процесса различения d_i и d_j. Разность (d_j - d_i) называется *различительной разностью*. При большом числе предъявлений двух стимулов различительные разности также формируют свое нормальное распределение на психологическом континууме. Поэтому среднее распределение разностей различения (d_j - d_i) будет равно разности средних распределений самих процессов различения — (S_j - S_i), а дисперсия распределения различительных разностей равна

$$s(d_j - d_i) = (s^2_j + s_i - 2r_{i,j}s_is_j)^{1/2}, \quad (1)$$

где s_i и s_j — дисперсии процессов различения i-го и j-го стимулов, соответственно, а $r_{i,j}$ — есть корреляция между мгновенными значениями процессов различения стимулов i и j.

Рассмотрим теперь следующую ситуацию. Пусть наблюдателю предъявляются пары стимулов i и j и от него требуется осуществить суждение, какой из стимулов дальше отстоит от нуля на психологическом континууме (например, более тяжелый или более сложный, или более красивый и т.д.). На рис. 1 показаны гипотетические процессы различения стимулов i и j.

Предполагается, что если различительный процесс для стимула j окажется на психологическом континууме выше, чем для стимула i, т.е. если различительная разность (d_j - d_i) > 0, то последует суждение, что стимул j больше, чем стимул i. И соответственно при (d_j - d_i) < 0 — произойдет обратное суждение.

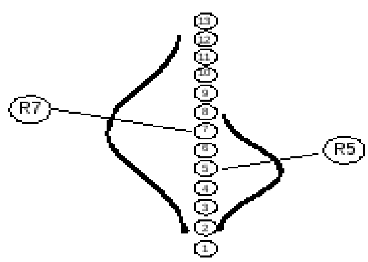

Рис.1. Гипотетическая модель процесса различения 2-х стимулов

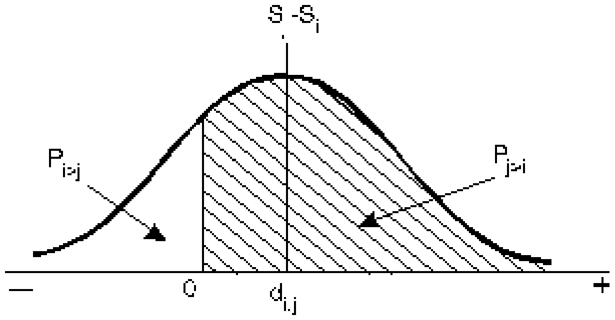

Рис. 2. Гипотетическое распределение процессов различения стимулов S_j и S_i на психологическом континууме: заштрихованная область указывает частоту суждения: стимул j больше, а незаштрихованная — стимул j меньше; d_{ij} - различие шкальных значений стимулов i и j, измеренное в единицах стандартного отклонения данного распределения — s (d_j-d_i)

Однако, если распределения различительных процессов перекрываются, то суждение, что стимул j меньше, чем стимул i может произойти даже тогда, когда величина S_j на психологическом континууме больше, чем величина S_i. На рис. 2 показано распределение различительных разностей при большом числе суждений.

Среднее распределения равно различию шкальных величин двух стимулов — (S_j - S_i). Это различие можно найти из таблицы областей под единичной нормальной кривой, зная пропорцию суждений стимул j больше, чем стимул i от общего числа суждений по данной паре стимулов (т.е., сделав стандартное преобразование “$p \rightarrow z$”).

В единицах дисперсии $s(d_j - d_i)$ это можно записать так:

$S_j - S_i = z_{j,i}s(d_j - d_i)$, (2)

где $z_{j,i}$ — обозначает искомое различие.

Подставляя это выражение в уравнение (1), получим:

$S_j - S_i = z_{j,i}(s_j^2 + s_i^2 - 2r_{i,j}s_is_j)^{1/2}$. (3)

Уравнение (3) и выражает в общем виде *закон сравнительных оценок Терстоун*