

Оценка соответствия. Аккредитация лабораторий

УДК 519.237+663.22

ПОЗИЦИОННЫЙ АНАЛИЗ КАК МЕТОД ОЦЕНКИ СОГЛАСОВАННОСТИ ЭКСПЕРТНЫХ ОЦЕНОК

© А. А. Халафян¹, З. А. Темердашев¹, Ю. Ф. Якуба²,
Н. В. Киселева¹, Т. И. Гугучкина², М. В. Антоненко²

Статья поступила 31 марта 2015 г.

В работе проанализированы различные методы оценки согласованности экспертных оценок, которые зависят от способов измерения свойств объектов. Показана возможность использования метода «Надежность и позиционный анализ» для оценки степени совпадения мнений экспертов в случае, когда результаты экспертных оценок представлены в балльной шкале. На примере дегустации вин показано, что данный метод многомерного анализа данных может быть использован для оценки согласованности экспертных оценок посредством вычисления статистики α Кронбаха как интегрального числового критерия надежности экспертизы. Отмечено, что оценка единодушия экспертов — важная проблема с учетом того, что экспертная оценка является достаточно субъективным процессом, зависящим не только от индивидуальных особенностей объектов экспертизы, но и от физического и психоэмоционального состояния экспертов, подверженных влиянию большого количества случайных неуправляемых факторов. Анализируемый метод позволяет выявить не только объекты экспертизы, по которым в наибольшей степени расходятся мнения экспертов, но и экспертов, мнения которых не согласуются с мнением большинства. Приведены результаты применения метода в среде пакета STATISTICA для оценки согласованности экспертов при проведении дегустации вин.

Ключевые слова: экспертная оценка; надежность и позиционный анализ; качество продукции; дегустация вин; статистика α Кронбаха.

Во многих областях знания и практической деятельности точное измерение переменных в принципе невозможно. В первую очередь это сложные процессы, не поддающиеся математической формализации, в таких областях, как экономика, управление, медицина, психология, социология, оценка вкусовых качеств пищевых продуктов и т.д. Тем не менее, в соответствии со словами Галилея: «Измеряй все, поддающееся измерению, и сделай таким все, не поддающееся измерению» проблема измерения свойств таких объектов сохраняет свою актуальность.

Одним из способов решения данной проблемы являются экспертные оценки — это оценки специалистами (экспертами) объектов или явлений в порядковой (балльной) шкале, если нет возможности произвести измерения посредством приборов. Так как экспертные оценки основаны на суждениях специалистов, то им свойственна определенная субъективность. Поэтому разрабатывают научные методы проведения экспертиз и обработки их результатов, чтобы

увеличить объективность, а значит и надежность экспертных оценок. Существуют различные способы формирования экспертных оценок. Наиболее часто применяют ранжирование, парное сравнение, непосредственную балльную оценку [1, 2].

Ранжирование — это расположение n объектов в порядке возрастания или убывания какого-либо свойства таким образом, чтобы объекту, обладающему наилучшим свойством, был приписан ранг 1, а объекту с наихудшим свойством — ранг n .

Парное сравнение — это установление предпочтительных 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. При использовании данного метода составляют матрицу размером $n \times n$, в ij -й ячейке которой проставляют 1, если объект i предпочтительнее объекта j , 0 — в противном случае, где i, j принимают значения 1, 2, ..., n . Тот объект, у которого сумма по строке окажется больше, и будет предпочтительным.

Балльная оценка предполагает разбиение условного диапазона изменения свойства объекта на отдельные интервалы, каждому из которых приписыв-

¹ Кубанский государственный университет, г. Краснодар, Россия; e-mail: statlab@kubsu.ru, temza@kubsu.ru

² Северо-Кавказский зональный научно-исследовательский институт садоводства и виноградарства, г. Краснодар, Россия; e-mail: globa2001@mail.ru

вают определенную оценку, например 0, 1, ..., 10 или 0, 1, ..., 100. Эксперт j ($j = 1, \dots, m$) в силу своих субъективных ощущений назначает объекту i ($i = 1, \dots, n$) тот или иной балл в зависимости от анализируемого свойства объекта. При этом разным объектам могут присваиваться одинаковые баллы. По сути таким методом оценивают знания учащихся. Предпочтительным будет тот объект, которому соответствует наибольший балл. Балльная оценка является наиболее информативным методом, так как позволяет не только упорядочить объекты по степени предпочтения, но и дает возможность оценить, насколько один объект предпочтительнее другого.

Как правило, экспертную оценку осуществляют несколько экспертов. Очевидно, что чем больше разброс их мнений, тем менее надежной будет экспертиза. Поэтому одним из показателей надежности экспертизы является единодушие или согласованность экспертов.

Существуют различные методы оценки согласованности экспертных оценок, которые зависят от способов измерения свойств объектов. При ранжировании объектов применяют коэффициент ранговой корреляции Спирмена (если сравнивают два объекта) или коэффициент конкордации Кендалла (если сравнивают более двух объектов). При использовании балльных оценок возможно применение таких статистик, как:

размах — разница между минимальной и максимальной оценками i -го объекта

$$R_i = (x_{\max})_i - (x_{\min})_i; \quad (1)$$

выборочное среднее квадратичное отклонение

$$S_i = \sqrt{S_i^2}, \quad (2)$$

где S_i^2 — выборочная дисперсия:

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2; \quad (3)$$

выборочный коэффициент вариации

$$V_i = \frac{S_i}{\bar{x}_i} \cdot 100\%. \quad (4)$$

Существенным недостатком приведенных статистик является то, что они принимают значения в неограниченном диапазоне, и сказать, что приемлемое единодушие экспертов, а значит, и надежность экспертизы, достигнута, затруднительно. Для коэффициента вариации принято считать, что согласованность экспертов удовлетворительная, если все V_i меньше 30, и хорошая, если все V_i меньше 20 [3]. Другим недостатком указанных статистик является то, что количество значений каждой из них равно числу объектов экспертизы (n), т.е. ни одна из них не является интегральным

показателем, характеризующим качество оценки объектов в совокупности.

Можно перевести оценки из балльной шкалы в ранговую с последующим вычислением коэффициента конкордации Кендалла. Но переход к более «бедной» шкале сопровождается значительной потерей информации: например, если трем объектам экспертом выставлены баллы 81, 85, 88, то в ранговой шкале этим баллам будут соответствовать значения 1, 2, 3, тем самым произойдет нивелирование разности баллов в оценке объектов.

Но по своей природе балльная шкала является порядковой шкалой, в которой выражается предпочтение одного объекта перед другим посредством баллов. В порядковой шкале не определена операция сложения, а значит, не определены и другие арифметические операции, тогда лишено смысла вычисление вышеупомянутых статистик. В настоящее время интенсивно разрабатывают математические методы статистики объектов нечисловой природы, составной частью которой является теория измерений [4]. Авторами работы [5] проведен обзор публикаций, посвященных выбору методов усреднения статистических данных в соответствии с принципами теории измерений. Н. А. Цейтлиным [6] предложен среднемедианный показатель в качестве оценки среднего для экспертных оценок короткого ряда (например, 5-балльная шкала оценки знаний учащихся). Проблеме построения интегральных оценок качества объектов по результатам экспертных оценок с учителем и без учителя посвящена работа [7]. А. И. Орловым с соавторами проведен обзор новых математических методов экспертных технологий, в частности, статистики объектов нечисловой природы [8], предложена технология экспертного оценивания условных вероятностей редких промежуточных событий, влекущих появление нежелательного события.

С другой стороны, большинство математических методов, как, впрочем, и математические модели, основаны на определенных допущениях или предположениях. Это и предположение о линейности, которое лежит в основе линейного программирования, и предположение о нормальности, являющееся основополагающим в параметрических методах статистики, и др. Неслучайно большинство теорем в высшей математике начинается со слов «предположим», «допустим» и т.д. В контексте сказанного для длинных дифференцированных рядов, например, 10-балльной или тем более 100-балльной шкалы, справедливым может быть допущение о числовой природе баллов. При таком предположении мы считаем баллы целыми числами, и тогда будет обоснованным применение методов математической статистики, построенных на парадигме среднего арифметического выборки.

В основе метода «Надежность и позиционный анализ» [9], который был разработан для построения

надежных анкет (шкал) в психологии, а также для анализа и улучшения используемой шкалы, лежит именно такое допущение. Рассмотрим применение данного метода, в частности, статистики α Кронбаха, для оценки надежности (согласованности) экспертных оценок. Оценивание надежности шкалы основано на корреляциях между индивидуальными позициями или измерениями, составляющими шкалу, и дисперсиями этих измерений. Результаты каждого измерения при заполнении анкеты (баллы, описывающие отношение респондента к данной позиции анкеты) включают в себя как истинное значение, так и частично неконтролируемую случайную погрешность. Результат измерения, т.е. отклик респондента на утверждение анкеты можно описать следующим уравнением:

$$X = \alpha + \delta,$$

где α обозначает неизвестное истинное значение, а δ — случайную ошибку, или погрешность измерения. Приведенное уравнение позволяет формализовать определение надежности: измерение является надежным, если его основную часть по отношению к погрешности составляет истинное значение. Поэтому в качестве статистики или критерия надежности утверждения (позиции анкеты) применяют отношение вариации (изменчивости) истинного значения, присущей респондентам, к общей вариации измерения:

$$\eta = \frac{\sigma_a^2}{\sigma_x^2},$$

где η — надежность; σ_a^2 — дисперсия истинного значения; σ_x^2 — дисперсия реального измерения.

Если ошибочная компонента в ответах респондентов на каждый вопрос действительно случайна, можно ожидать, что в ответах на различные вопросы случайные компоненты будут взаимно подавлять друг друга. Тогда если просуммировать несколько измерений — ответов респондента, описывающих его отношение к проблеме, — суммарная погрешность по совокупности всех утверждений анкеты будет близка к нулю. При этом среднее арифметическое совокупности реальных измерений (x) будет приблизительно равно среднему арифметическому истинных значений (a)

$$\bar{x} = \frac{1}{m} \sum_{j=1}^m a_j + \frac{1}{m} \sum_{j=1}^m \delta_j \approx \frac{1}{m} \sum_{j=1}^m a_j + \frac{1}{m} 0 \approx \bar{a}. \quad (5)$$

Мы воспользовались тем свойством, что сумма погрешностей примет значение, близкое к 0, т.е.

$$\sum_{j=1}^n \delta_j \approx 0, \text{ причем чем больше количество утверждений (позиций анкеты) } n, \text{ тем ближе к 0 будет сумма,}$$

а значит, более верным будет равенство средних арифметических реальных измерений и истинных значений. Следовательно, чем больше будет утверждений

в анкете, тем точнее истинное значение будет отражено на суммарной шкале, полученной суммированием баллов ответов всех респондентов. Этот вывод описывает важный принцип создания анкет, а именно: чем больше утверждений участвует в построении анкеты для оценки данной концепции, тем более надежным будет общее измерение — суммарная шкала.

В качестве общепринятого индекса надежности анкеты применяют статистику α Кронбаха:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n s_i^2}{s_{\text{сум}}^2} \right], \quad (6)$$

где s_i^2 — выборочная дисперсия i -го утверждения; $s_{\text{сум}}^2$ — выборочная дисперсия суммарной шкалы (суммарной оценки респондентов).

Эта статистика позволяет оценить долю дисперсии (изменчивости) истинного значения по совокупности всех утверждений анкеты — суммарной шкалы, путем сравнения суммы дисперсий отдельных утверждений $\sum s_i^2$ с дисперсией суммарной шкалы $s_{\text{сум}}^2$.

Если не существует истинного значения, а присутствует только случайная погрешность в оценке респондентами своего отношения к утверждениям анкеты, то дисперсия суммарной шкалы будет такой же, как сумма дисперсий отдельных утверждений (отдельного эксперта), т.е. $\sum s_i^2 = s_{\text{сум}}^2$, в этом случае α будет равна 0. Если все утверждения совершенно надежны и измеряют одно и то же истинное значение, то α будет равна 1, так как $\sum s_i^2 / s_{\text{сум}}^2 = 1/n$. Естественно, в реальных случаях можно говорить лишь о близости статистики α к 0 или 1.

Как было ранее сказано, анкета тем более надежна, чем больше позиций она содержит. В модуле «Надежность и позиционный анализ», реализованном в программе STATISTICA [8], предусмотрена возможность выбора оптимального состава и количества позиций, повышающих надежность шкалы. Программа позволяет проводить анализ шкалы, содержащей до 300 позиций. Посредством определенных статистик можно удалить из шкалы неинформативные, избыточные позиции. Из шкалы желательно исключить позиции, имеющие отрицательные корреляции с суммарной шкалой, позиции с малыми значениями коэффициентов корреляции и детерминации, а также те, которым соответствуют значения α , большие, чем α Кронбаха для всей шкалы. Манипулируя позициями в интерактивном режиме (удаляя избыточные позиции и добавляя новые), можно построить максимально надежную шкалу.

Так как согласованность мнений экспертов (единодушие) также зависит от величины разброса баллов при оценке объектов, ничто не мешает использовать статистику α Кронбаха в качестве интегрального пока-

Таблица 1. Результаты оценки испытуемых образцов вин по 100-балльной шкале

№ экс- перта	Образцы вин														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	85	80	83	85	88	81	84	84	87	87	80	85	83	82	85
2	82	80	80	84	83	88	88	88	80	83	84	86	88	83	82
3	82	84	79	80	82	87	88	87	78	81	83	84	87	81	82
4	87	83	79	84	81	84	85	84	82	82	80	84	90	83	81
5	84	80	80	81	80	81	83	83	84	86	82	86	81	80	82
6	85	81	82	80	80	83	80	85	82	84	85	85	77	82	80
7	84	81	82	81	79	80	83	86	87	82	84	82	78	81	81
8	83	80	81	81	79	78	81	83	85	83	78	81	78	85	80
9	86	71	98	98	75	88	89	84	88	82	80	97	98	89	98
10	83	81	83	84	83	77	83	84	86	80	79	85	80	84	81
11	85	80	83	85	81	69	66	85	76	73	71	81	60	68	64
12	89	80	82	100	87	82	83	85	88	80	70	85	91	81	63
13	85	80	83	85	81	81	83	84	84	82	80	88	86	89	86

№ экс- перта	Образцы вин														Сумма
	16	17	18	19	20	21	22	23	24	25	26	27	28		
1	88	84	84	83	88	84	87	87	86	84	86	90	90	2380	
2	88	87	84	83	89	80	84	89	87	82	88	90	90	2380	
3	88	88	84	78	89	80	80	89	86	79	89	92	89	2356	
4	91	83	80	79	91	84	81	83	82	80	90	91	91	2355	
5	85	82	80	80	90	82	80	82	81	78	90	90	80	2313	
6	75	86	77	78	95	81	84	80	83	80	90	92	93	2325	
7	83	84	81	80	86	82	82	83	84	82	85	86	86	2315	
8	82	80	85	78	85	80	79	79	83	80	84	86	87	2284	
9	90	86	86	78	98	88	86	86	86	86	96	86	98	2471	
10	84	89	86	82	85	85	98	91	87	86	99	94	97	2396	
11	69	65	81	72	71	61	74	68	71	64	93	86	72	2074	
12	82	84	77	73	84	75	80	78	87	71	87	85	90	2299	
13	90	92	90	90	90	88	92	88	94	90	94	94	94	2443	

зателя надежности экспертизы. Применительно к экспертным оценкам в качестве респондентов целесообразно рассматривать экспертов, проводящих экспертизу, а в качестве позиций анкеты — объекты, подвергнутые экспертизе. Тогда чем больше объектов участвует в экспертной оценке, тем выше согласованность мнений экспертов.

Обработка экспертных оценок — это проверка согласованности мнений экспертов (или классификация экспертов, если нет согласованности) и усреднение мнений экспертов внутри согласованной группы. Дегустация пищевых продуктов является основным способом оценки вкусовых качеств испытуемых образцов посредством анализа их органолептических свойств [10, 11].

При проведении органолептического анализа вин результаты дегустации могут быть обработаны с помощью критерия χ^2 , при малом числе ответов можно использовать критерий Фишера, а для определения порога восприятия — теорию обнаружения импульса [12]. Установить различие или подобие между двумя однородными продуктами по интенсивности исследуемой характеристики позволяет метод парного сравнения [13]. Однако индивидуальная чувствительность испытателей вносит большую долю субъективности в оценку результатов органолептических различитель-

ных испытаний. В этом случае не представляется возможным сделать общее заключение относительно воспроизводимости результатов, применимое к любой группе испытателей [14].

Рассмотрим оценку согласованности (единодушия) экспертов при дегустации вин, проведенной в федеральном государственном научном учреждении «Северо-Кавказский зональный научно-исследовательский институт садоводства и виноградарства» (Краснодар). Для оценки согласованности экспертных оценок применен метод многомерного анализа данных с вычислением статистики α Кронбаха как интегрального числового критерия надежности экспертизы. В качестве исходных данных использовали результаты дегустационных испытаний (табл. 1). Тринадцать экспертов оценили 28 образцов вин по 100-балльной шкале. В последнем столбце таблицы отображена сумма баллов, выставленных каждым экспертом, называемая в терминологии позиционного анализа суммарной шкалой. В этой терминологии таблица — это анкета, строки, обозначающие экспертов, — респонденты, а столбцы, обозначающие образцы вин, — позиции (утверждения) анкеты.

Перед тем как перейти к анализу надежности, исследуем описательные статистики измерений. В табл. 2 отображены средние, стандартные отклоне-

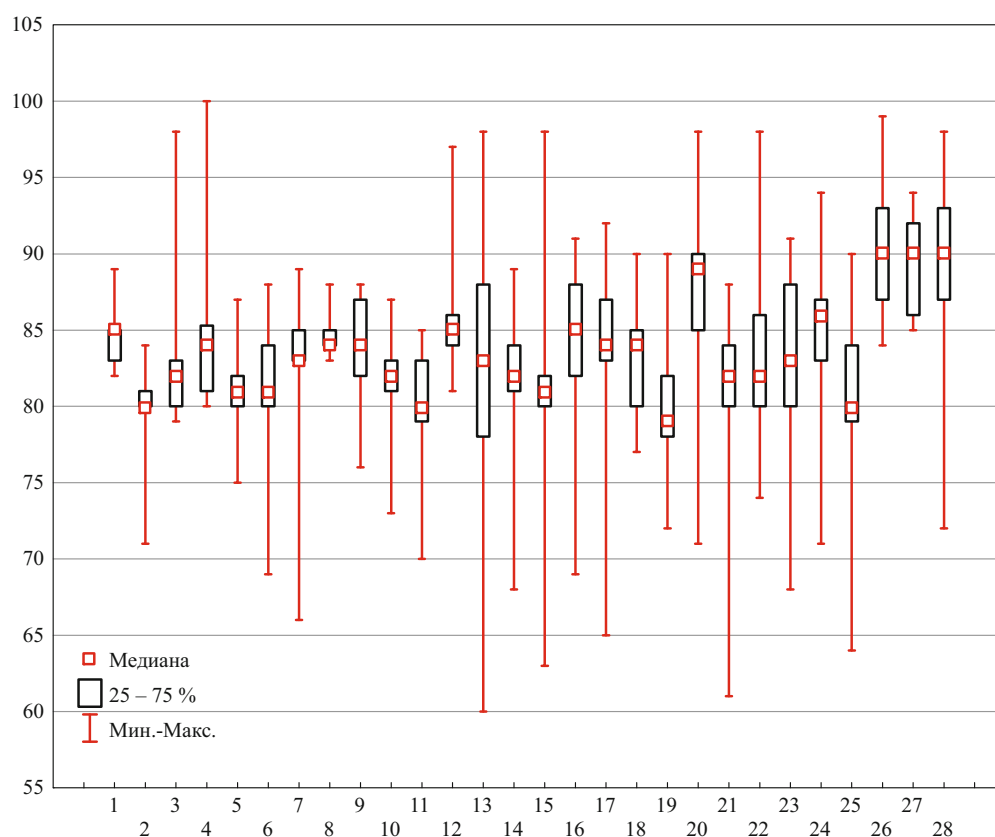


Рис. 1. Диаграммы размаха экспертных оценок качества вин

ния, дисперсии, коэффициенты вариации для всех 28 образцов. Как видно из таблицы, средние значения оценок всех экспертов принадлежат достаточно широкому диапазону (от 80 до 90), что свидетельствует о высокой изменчивости мнений экспертов. Наибольший средний балл у образца 26 ($\bar{x} = 90,08$), наименьший — у образца 19 ($\bar{x} = 79,54$). Размах $R = 10,54 \approx 11$. При этом образцам 13, 15, 28 соответствуют наибольшие разбросы, стандартные отклонения равны 9,2; 8,87; 6,93 соответственно. Наименьшие разбросы у образцов 8, 1, 2 — стандартные отклонения составляют 1,98; 3,01; 4,84 соответственно.

На рис. 1 приведены диаграммы размаха экспертных оценок: для каждого образца можно визуально оценить степень симметричности распределений дегустационных баллов, среднее (медиану), изменчивость (квартильный размах — расстояние между 75-м и 25-м процентилями), размах (расстояние между наибольшим и наименьшим значениями). Квартильный размах и размах характеризуют разброс дегустационных оценок образцов вин.

Из графика видно, что распределения баллов преимущественно не симметричны, исключением является образец 15. Этот результат легко объяснить отклонением эмпирических распределений дегустационных баллов от нормального закона. Наибольшая медиана, равная 90 (медиана применяется для оценки среднего при несимметричных распределениях), у последних трех образцов 26, 27, 28 свидетельствует о высокой оценке их качества дегустаторами. Наимень-

Таблица 2. Описательные статистики экспертных оценок качества вин

Образец	Среднее $\bar{x}$	Стандартное отклонение s_i	Дисперсия s_i^2	Коэффициент вариации $V_i, \%$
1	84,62	1,98	3,92	2,34
2	80,08	3,01	9,08	3,76
3	82,69	4,84	23,40	5,85
4	85,23	6,42	41,19	7,53
5	81,46	3,38	11,44	4,15
6	81,46	5,13	26,27	6,29
7	82,77	5,72	32,69	6,91
8	84,77	1,48	2,19	1,75
9	83,62	3,84	14,76	4,59
10	81,92	3,38	11,41	4,12
11	79,69	4,61	21,23	5,78
12	85,31	4,05	16,40	4,75
13	82,85	9,20	84,64	11,10
14	82,15	5,13	26,31	6,24
15	80,38	8,87	78,59	11,03
16	84,23	6,35	40,36	7,54
17	83,85	6,48	41,97	7,73
18	82,69	3,77	14,23	4,56
19	79,54	4,56	20,77	5,73
20	87,77	6,42	41,19	7,31
21	80,77	6,91	47,69	8,55
22	83,62	6,17	38,09	7,38
23	83,31	6,20	38,40	7,44
24	84,38	5,17	26,76	6,13
25	80,15	6,69	44,81	8,35
26	90,08	4,41	19,41	4,89
27	89,38	3,23	10,42	3,61
28	89,00	6,93	48,00	7,78

Таблица 3. Итоги позиционного анализа дегустационной оценки качества вин

Образец	Среднее при удалении	Дисперсия при удалении	Стандартное отклонение при удалении	Корреляция между позицией и суммарной шкалой	α при удалении
1	2253,15	8613,98	92,81	-0,08	0,9459
2	2257,69	8760,52	93,60	-0,33	0,9480
3	2255,08	8306,38	91,14	0,31	0,9447
4	2252,54	8411,33	91,71	0,12	0,9479
5	2256,31	8651,14	93,01	-0,12	0,9471
6	2256,31	7931,29	89,06	0,72	0,9406
7	2255,00	7726,92	87,90	0,86	0,9389
8	2253,00	8601,39	92,74	-0,05	0,9456
9	2254,15	8246,75	90,81	0,49	0,9430
10	2255,85	8232,90	90,74	0,59	0,9425
11	2258,08	8174,53	90,41	0,49	0,9430
12	2252,46	8107,79	90,04	0,67	0,9416
13	2254,92	7313,15	85,52	0,79	0,9400
14	2255,61	7773,16	88,17	0,91	0,9387
15	2257,39	7394,55	85,99	0,77	0,9403
16	2253,54	7671,17	87,59	0,82	0,9391
17	2253,92	7577,92	87,05	0,90	0,9381
18	2255,08	8263,30	90,90	0,48	0,9432
19	2258,23	8053,10	89,74	0,66	0,9415
20	2250,00	7676,46	87,62	0,81	0,9393
21	2257,00	7475,23	86,46	0,93	0,9374
22	2254,15	7835,98	88,52	0,69	0,9409
23	2254,46	7648,56	87,46	0,87	0,9385
24	2253,39	7821,16	88,44	0,85	0,9393
25	2257,61	7553,31	86,91	0,89	0,9381
26	2247,69	8418,37	91,75	0,20	0,9455
27	2248,39	8337,77	91,31	0,43	0,9436
28	2248,77	7541,87	86,84	0,87	0,9384

Примечание. Итог для шкал: среднее значение — 2337,77; стандартное отклонение — 96,469; количество экспертов — 13; α Кронбаха — 0,944; стандартизованное значение $\alpha = 0,928$.

шая медиана у образца 19, которому также соответствует наименьшее значение среднего. При этом следует обратить внимание на то, что относительно небольшие разбросы в оценке качества у образцов вин 8, 1, 27, 2, 5, 9, 10, 18 и достаточно большие у образцов 13, 15, 25, 28, 17.

Все, что было рассмотрено выше, относится к исследованию описательных статистик, позволяющих получить общие характеристики результатов дегустации. Для анализа согласованности мнений дегустаторов надо воспользоваться табл. 3, в которой отображены основные статистики суммарной шкалы (суммы баллов, проставленных экспертами для каждого образца вин, см. последний столбец табл. 1). Среднее значение суммарной шкалы равно 2337,77, стандартное отклонение — 96,47. В табл. 3 приведены две оценки надежности суммарной шкалы — α Кронбаха для балльной (0,94) и стандартизованной (0,93) шкал, последняя получается, если применить нормирование измерений. Нормирование переменной заключается в вычитании из каждого значения среднего значения и делении результата на стандартное отклонение. Нормирование нужно, если разные позиции предполагают разные системы балльной оценки. Оба значения α Кронбаха близки к 1, что говорит о до-

статочно высоком единодушии экспертов, а значит, и о надежности экспертных оценок.

Показатели, приведенные в первых трех столбцах табл. 3, позволяют ответить на вопрос, как изменятся основные статистики суммарной шкалы при удалении из результатов дегустации соответствующего образца вина. В четвертом столбце отображены множественные корреляции между оценками экспертов и суммарной шкалой, если удалить оценки данного образца. Наибольший интерес представляет последний столбец, в котором приведены значения α Кронбаха при удалении данного образца.

Образцы вин, которым соответствуют малые корреляции и значения α Кронбаха при удалении, превышающие α Кронбаха для суммарной шкалы (0,944), снижают надежность экспертизы. Из табл. 3 видно, что это образцы 1 – 5, 8, 26, так как значения α Кронбаха при удалении (0,946; 0,948, 0,945; 0,948; 0,947; 0,946; 0,946 соответственно) превышают 0,944. Корреляции, соответствующие этим образцам вин, также принимают малые значения. Указанные образцы в табл. 3 отмечены полужирным шрифтом.

Чем меньше α Кронбаха в последнем столбце, тем вклад образца в надежность экспертизы или в единодушие экспертов выше: наибольший вклад в надеж-

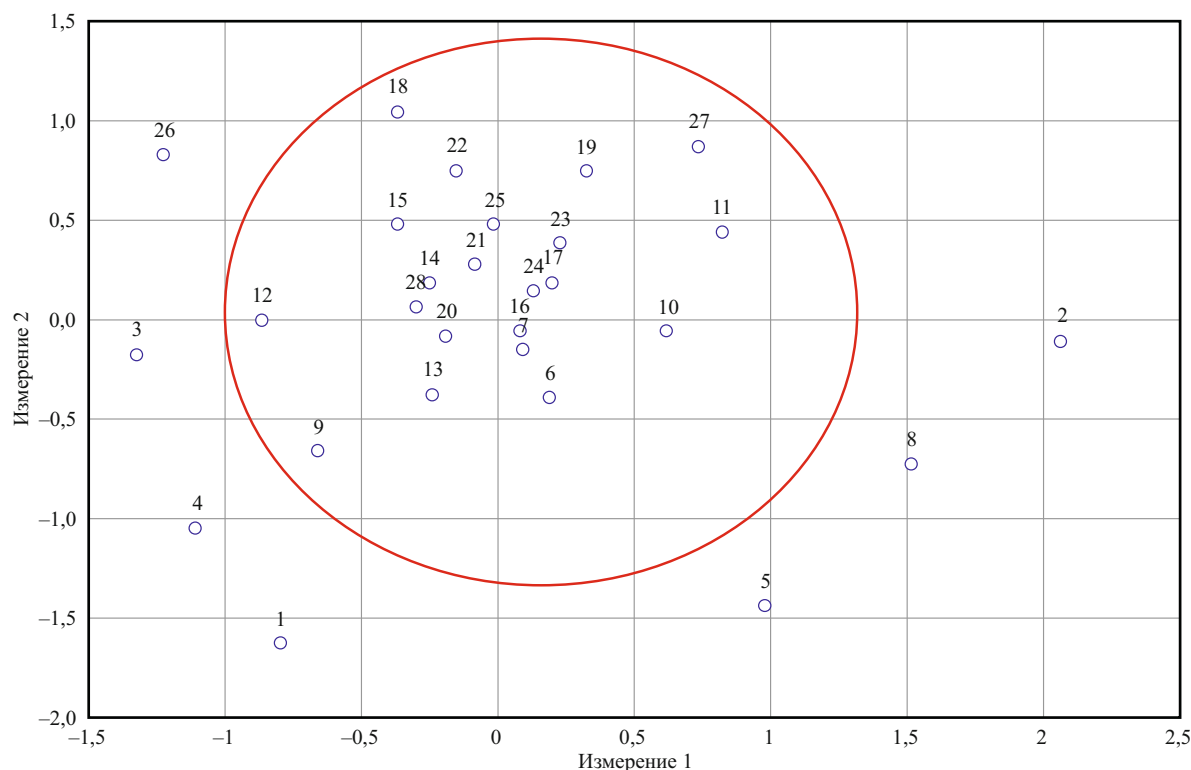


Рис. 2. Результаты многомерного шкалирования испытуемых образцов вин

ность дегустации вносят образцы 21, 25, 17, 28, 13, 14, 7, 16, 24.

Многомерное шкалирование позволяет перенести объекты многомерного пространства в пространство меньшей размерности, в частности, на плоскость, сохранив порядок расстояний между объектами. Объекты, которые близки в многомерном пространстве, сохраняют близость и в пространстве меньшей размерности.

Если воспользоваться матричным файлом таблицы корреляций, создание и сохранение которого предусмотрены в модуле «Надежность и позиционный анализ», то при помощи многомерного шкалирования образцы вин, являющиеся объектами в пространстве размерности 13, можно представить в виде точек на плоскости, максимально сохранив порядок парного сходства между ними относительно экспертных оценок. Чем выше сходство между образцами вин, тем меньше будет расстояние между ними.

На рис. 2, а образцы вин представлены в виде точек на плоскости в системе координат измерение 1 — измерение 2.

Следует обратить внимание, что образцы вин, для которых α Кронбаха при удалении меньше α Кронбаха для суммарной шкалы, образуют единый кластер, обозначенный на графике эллипсом. Образцы вин, принадлежащие данному кластеру, вносят положительный вклад в надежность экспертизы, увеличивая единодушие дегустаторов. Образцы 1 – 5, 8, 26, для которых α Кронбаха при удалении превышает α Кронбаха для суммарной шкалы, расположены на

периферии диаграммы за пределами выделенного кластера. Эти образцы вносят отрицательный вклад в надежность дегустации, уменьшая согласованность экспертов.

В модуле «Надежность и позиционный анализ» предусмотрена возможность определения необходимого числа позиций, чтобы при сохранении основных статистических показателей суммарной шкалы была достигнута требуемая надежность анкеты. Например, если выбрать уровень надежности экспертизы (α Кронбаха), равный 0,95, то для его достижения надо добавить еще четыре образца, а в случае уровня надежности 0,98 — еще 54 образца вин.

Если оценивать надежность экспертизы по совокупности коэффициентов вариации для каждого образца вин, представленных в последнем столбце табл. 2, то согласованность экспертов хорошая, так как все коэффициенты вариации V_i меньше 20 [2].

Показатель надежности экспертных оценок α Кронбаха можно вычислить «вручную», не прибегая к модулю «Надежность и позиционный анализ». Для этого достаточно воспользоваться формулой (6), если учесть, что дисперсия суммарной шкалы, вычисленная по формуле (3), равна 96,47, $n = 28$, а дисперсии s_i^2 отображены во втором столбце табл. 2:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{1}{s_{\text{сум}}^2} \sum s_i^2 \right] = \frac{28}{27} \left[1 - \frac{1}{(96,47)^2} 835,61 \right] = 0,944.$$

Таким образом, мы рассмотрели различные способы вычисления статистики α Кронбаха, которая

Таблица 4. Транспонированная таблица результатов дегустации

№ образца	Эксперт												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	85	82	82	87	84	85	84	83	86	83	85	89	85
2	80	80	84	83	80	81	81	80	71	81	80	80	80
3	83	80	79	79	80	82	82	81	98	83	83	82	83
4	85	84	80	84	81	80	81	81	98	84	85	100	85
5	81	83	82	81	80	80	79	79	75	83	81	87	81
6	81	88	87	84	81	83	80	78	88	77	69	82	81
7	84	88	88	85	83	80	83	81	89	83	66	83	83
8	84	88	87	84	83	85	86	83	84	84	85	85	85
9	87	80	78	82	84	82	87	85	88	86	76	88	84
10	87	83	81	82	86	84	82	83	82	80	73	80	82
11	80	84	83	80	82	85	84	78	80	79	71	70	80
12	85	86	84	84	86	85	82	81	97	85	81	85	88
13	83	88	87	90	81	77	78	78	98	80	60	91	86
14	82	83	81	83	80	82	81	85	89	84	68	81	89
15	85	82	82	81	82	80	81	80	98	81	64	63	86
16	88	88	88	91	85	75	83	82	90	84	69	82	90
17	84	87	88	83	82	86	84	80	86	89	65	84	92
18	84	84	84	80	80	77	81	85	86	86	81	77	90
19	83	83	78	79	80	78	80	78	78	82	72	73	90
20	88	89	89	91	90	95	86	85	98	85	71	84	90
21	84	80	80	84	82	81	82	80	88	85	61	75	88
22	87	84	80	81	80	84	82	79	86	98	74	80	92
23	87	89	89	83	82	80	83	79	86	91	68	78	88
24	86	87	86	82	81	83	84	83	86	87	71	87	94
25	84	82	79	80	78	80	82	80	86	86	64	71	90
26	86	88	89	90	90	90	85	84	96	99	93	87	94
27	90	90	92	91	90	92	86	86	86	94	86	85	94
28	90	90	89	91	80	93	86	87	98	97	72	90	94

может быть использована в качестве интегрального показателя надежности дегустации, так как характеризует согласованность экспертов посредством оценки изменчивости органолептических свойств объектов дегустации.

Однако необходимо учитывать тот факт, что в случае экспертных оценок надежность экспертизы зависит не только от изменчивости оцениваемых объектов (в нашем случае — образцов вин), но и от изменчивости мнений экспертов ввиду их определенной субъективности. Это означает, что на согласованность мнений экспертов влияют не только индивидуальные особенности объектов экспертизы (чем более они различны, тем менее единодушны эксперты), но и индивидуальные особенности экспертов в восприятии органолептических свойств объектов (чем более они различны, тем менее единодушны эксперты). Модуль «Надежность и позиционный анализ» позволяет оценить надежность экспертизы посредством оценки вариативности мнений экспертов. Для этого достаточно транспонировать табл. 1, заменив строки столбцами. В терминологии позиционного анализа это означает, что строки, обозначающие вина, — это респонденты, а столбцы, обозначающие экспертов, — позиции (утверждения) анкеты. Тогда результаты дегустации можно представить в виде табл. 4.

Перед тем как перейти к анализу надежности, исследуем описательные статистики измерений. В табл. 5 приведены средние и стандартные откло-

нения для всех 13 экспертов: средние значения оценок принадлежат достаточно широкому диапазону от 70 до 90, что свидетельствует о высокой изменчивости мнений экспертов. Наибольший средний балл у эксперта 9 ($\bar{x} = 88,23$), наименьший — у эксперта 11 ($\bar{x} = 74,07$). Размах $R = 14,6 \approx 15$. При этом экспертам 9, 11, 12 соответствуют наибольшие разбросы, стандартные отклонения равны 7,36; 8,61; 7,39 соответственно. Эти результаты уже говорят о некоторой несогласованности мнений экспертов, причем эта несогласованность вызвана в первую очередь экспертами 9 и 11. В какой степени выражена эта несогласованность и как она влияет на надежность дегустации, покажет дальнейший более углубленный позиционный анализ.

На рис. 3 приведены диаграммы размаха дегустационных оценок для каждого эксперта: можно визуально оценить степень симметричности распределений дегустационных баллов, медиану, квартильный размах, размах. Из графика видно, что распределения баллов преимущественно не симметричны, исключением являются эксперты 3 и 12.

Наибольшая медиана у дегустатора 9, что свидетельствует о его наиболее лояльном отношении к качеству образцов вин. Наименьшая медиана у дегустатора 11, что говорит о его наиболее критическом отношении к качеству образцов. При этом следует обратить внимание на то, что разбросы в оценке качества

Таблица 5. Описательные статистики оценок экспертов

Эксперимент	Среднее значение оценок	Стандартное отклонение
1	84,733	2,712
2	85,004	3,281
3	84,142	4,052
4	84,107	3,937
5	82,607	3,247
6	83,026	4,811
7	82,678	2,294
8	81,571	2,686
9	88,232	7,361
10	85,571	5,613
11	74,068	8,614
12	82,111	7,389
13	87,232	4,560

вин относительно невелики у экспертов 1 – 5, 7, 8, 13 и значительны у экспертов 6, 9 – 12.

Итоги позиционного анализа представлены в табл. 6. Среднее значение суммарной шкалы равно 1085,09, стандартное отклонение — 37,58. В табл. 6 приведены две оценки надежности суммарной шкалы — α Кронбаха для балльной (0,829) и стандартизованной (0,88) шкал. Оба значения α Кронбаха близки к 1, что говорит о надежности экспертных оценок. Но следует обратить внимание, что значение α Кронбаха (0,829), вычисленное посредством анализа изменчивости мнений каждого дегустатора при оценке всех образцов вин, значительно меньше ранее вычисленного посредством анализа изменчивости мнений всех дегустаторов при оценке каждого образца вина (0,944). Это означает, что согласованность дегустаторов при оценке их индивидуальных отличий в восприятии органолептических свойств вина меньше, чем согласованность при оценке индивидуальных свойств образцов вин.

Для того, чтобы выявить экспертов, снижающих единодушные дегустации, обратимся к последнему

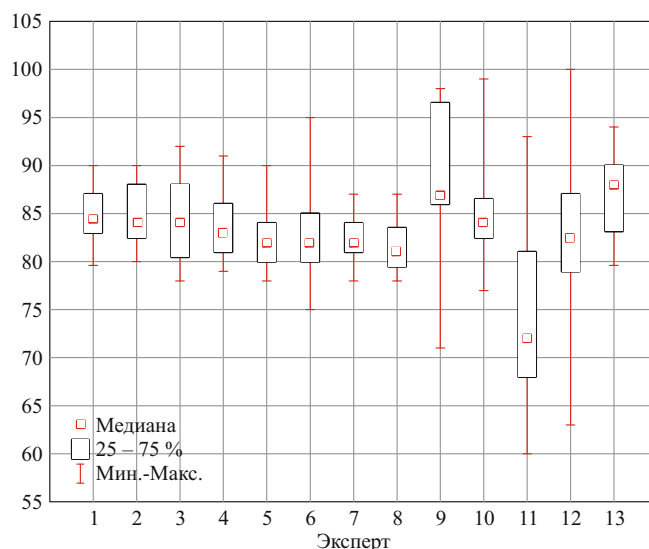


Рис. 3. Диаграммы размаха дегустационных оценок для каждого эксперта

столбцу, в котором приведены значения α Кронбаха при удалении данного эксперта. Из таблицы видно, что это эксперты 11 и 9 снижают согласованность дегустации, так как значения α Кронбаха при удалении (0,851 и 0,834 соответственно) превышают значение α Кронбаха для суммарной шкалы, равное 0,829. Корреляции также принимают малые значения — 0,281 и 0,352.

Чем меньше α Кронбаха в последнем столбце, тем вклад эксперта в согласованность экспертизы выше. Поэтому справедливым будет утверждение, что наиболее единодушными являются эксперты 4, 6, 10, 1, 5, 3, 2, 8, 13, 7 (номера экспертов указаны в порядке убывания их вклада в надежность).

На рис. 4 эксперты представлены в виде точек на плоскости в системе координат измерение 1 — измерение 2. Из графика видно, что эксперты 2 – 8, 10, 13 образуют кластер сходства при оценке качества вин

Таблица 6. Итоги позиционного анализа оценок для каждого эксперта

Эксперт	Среднее при удалении	Дисперсия при удалении	Стандартное отклонение при удалении	Корреляция между позицией и суммарной шкалой	α при удалении
1	1000,355	1225,866	35,012	0,691	0,812
2	1000,083	1219,026	34,914	0,589	0,814
3	1000,945	1195,102	34,570	0,548	0,813
4	1000,981	1161,444	34,079	0,704	0,804
5	1002,481	1215,548	34,864	0,612	0,813
6	1002,061	1136,810	33,716	0,636	0,805
7	1002,409	1260,615	35,505	0,601	0,818
8	1003,516	1234,307	35,132	0,651	0,814
9	996,855	1137,749	33,730	0,352	0,834
10	999,516	1110,985	33,331	0,600	0,806
11	1011,019	1130,473	33,622	0,281	0,851
12	1002,976	1088,342	32,990	0,461	0,823
13	997,856	1185,146	34,425	0,508	0,815

Примечание. Итог для шкалы: среднее — 1085,09; стандартное отклонение — 37,58; число испытуемых образцов — 28; α Кронбаха — 0,829042; стандартизованное значение α — 0,888280; средняя межпозиционная корреляция — 0,398979.

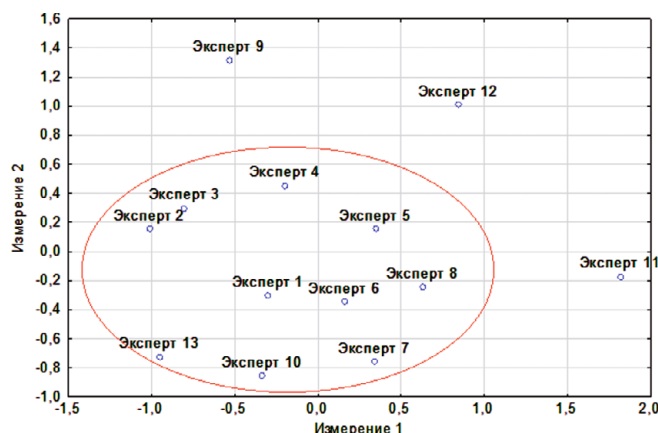


Рис. 4. Результаты многомерного шкалирования экспертов

(расположены внутри эллипса). Эксперты 9, 11, 12 образуют самостоятельные кластеры, состоящие из одного элемента. Таким образом, наибольшие отличия в оценке образцов вин у экспертов 9, 11, 12.

Проведенное исследование выявило дегустаторов, мнения которых оказались недостаточно согласованными с мнениями большинства экспертного сообщества. Это не означает, что данные дегустаторы — более (менее) квалифицированные эксперты, а означает лишь то, что их мнение отличается от мнения большинства.

Таким образом, в работе на примере дегустации вин показано, что метод многомерного анализа данных «надежность и позиционный анализ» может быть использован для оценки согласованности экспертных оценок посредством вычисления статистики α Кронбаха как интегрального числового критерия надежности экспертизы. Оценка единодушия экспертов — важная проблема, так как экспертная оценка является достаточно субъективным процессом, зависящим не только от индивидуальных особенностей объектов экспертизы, но и от физического и психоэмоционального состояния экспертов, подверженных влиянию большого количества случайных неуправляемых факторов. Метод позволяет выявить не только объекты экспертизы, по которым в наибольшей степени расходятся мнения экспертов, но и экспертов, мнения которых не согласуются с мнением большинства.

ЛИТЕРАТУРА

1. Орлов А. И. Организационно-экономическое моделирование. Ч. 2. Экспертные оценки. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. — 486 с.
2. Орлов А. И. Теория экспертных оценок в нашей стране / Политетический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 93. С. 1652 – 1683.
3. Григан А. М. Управленческая диагностика: теория и практика. — Ростов-на-Дону: Изд-во РСЭИ, 2009. — 316 с.
4. Орлов А. И. Математические методы исследования в теории измерений / Заводская лаборатория. Диагностика материалов. 2006. Т. 72. № 1. С. 67 – 70.
5. Барский Б. В., Соколов М. В. Средние величины, инвариантные относительно допустимых преобразований шкал измерения / Заводская лаборатория. Диагностика материалов. 2006. Т. 72. № 1. С. 59 – 66.

6. Цейтлин Н. А. Среднемедианный показатель положения выборки экспертных оценок / Заводская лаборатория. Диагностика материалов. 2010. Т. 76. № 7. С. 69 – 72.
7. Стрижов В. В. Уточнение экспертных оценок, выставленных в ранговых шкалах, с помощью измеряемых данных / Заводская лаборатория. Диагностика материалов. 2011. Т. 77. № 7. С. 72 – 78.
8. Орлов А. И., Савинов Ю. Г., Богданов А. Ю. Экспертные технологии и их применение при оценивании вероятностей редких событий / Заводская лаборатория. Диагностика материалов. 2014. Т. 80. № 3. С. 63 – 68.
9. Надежность и позиционный анализ. URL: <http://statsoft.ru/home/textbook/modules/streliab.html#general>.
10. Валушко Г. Г., Шольц-Куликов Е. П. Теория и практика дегустации вина. — Симферополь: Таврида, 2005. — 232 с.
11. The Aroma of Rojal Red Wines from La Mancha Region — Determination of Key Odorants / Ed. by Dr. Bekir Salih. — In Tech Europe, Rijeka, Croatia. P. 147 – 170.
12. ГОСТ Р ИСО 8588–2008. Органолептический анализ. Методология. Испытания «А» — «Не А». — М.: Стандартинформ, 2009. — 12 с.
13. ГОСТ Р 53161–2008. Органолептический анализ. Методология. Метод парного сравнения. — М.: Стандартинформ, 2009. — 20 с.
14. ГОСТ Р 53159–2008. Органолептический анализ. Методология. Метод треугольника. — М.: Стандартинформ, 2009. — 16 с.

REFERENCES

1. Orlov A. I. Organizatsionno-ekonomicheskoe modelirovanie [Organizational-economic modeling]. Part 2. Ékspertnye otsenki [Expert evaluation]. — Moscow: Izd. MG TU im. N. É. Bauman, 2011. — 486 p. [in Russian].
2. Orlov A. I. Teoriya ékspertnykh otsenok v nashei strane [The theory of expert assessments in our country] / Politem. Set. Élektron. Nauch. Zh. Kub. Gos. Agrar. Univ. 2013. N 93. P. 1652 – 1683 [in Russian].
3. Grigan A. M. Upravlencheskaya diagnostika: teoriya i praktika [Managerial diagnostics: theory and practice]. — Rostov-on-Don: Izd. RSÉI, 2009. — 316 p. [in Russian].
4. Orlov A. I. Matematicheskie metody issledovaniya v teorii izmerenii [Mathematical methods of research in the theory of measurement] / Zavod. Lab. Diagn. Mater. 2006. Vol. 72. N 1. P. 67 – 70 [in Russian].
5. Barskii B. V., Sokolov M. V. Srednie velichiny, invariantnye otноситel'no dopustimyykh preobrazovaniy shkal izmereniya [Mean values invariant under admissible transformations of scale of measurement] / Zavod. Lab. Diagn. Mater. 2006. Vol. 72. N 1. P. 59 – 66 [in Russian].
6. Tseitlin N. A. Srednemediannyi pokazatel' polozheniya vyborki ékspertnykh otsenok [Average median indicator of setting sampling of expert assessments] / Zavod. Lab. Diagn. Mater. 2010. Vol. 76. N 7. P. 69 – 72 [in Russian].
7. Strizhov V. V. Utochnenie ékspertnykh otsenok, vystavlenykh v rangovykh shkalakh, s pomoshch'yu izmeryaemykh dannykh [Refinement of expert estimates in the rank scale using measured data] / Zavod. Lab. Diagn. Mater. 2011. Vol. 77. N 7. P. 72 – 78 [in Russian].
8. Orlov A. I., Savinov Yu. G., Bogdanov A. Yu. Ékspertnye tekhnologii i ikh primeneniye pri otsenivaniy veroyatnostei redkikh sobyitiy [Expert system technology in estimating the probability of rare events] / Zavod. Lab. Diagn. Mater. 2014. Vol. 80. N 3. P. 63 – 68 [in Russian].
9. Nadezhnost' i pozitsionnyi analiz [Reliability and Item Analysis]. URL: <http://statsoft.ru/home/textbook/modules/streliab.html#general> [in Russian].
10. Valuiko G. G., Shol'ts-Kulikov E. P. Teoriya i praktika degustatsii vina [Theory and practice of wine tasting]. — Simferopol: Tavrida, 2005. — 232 p. [in Russian].
11. The Aroma of Rojal Red Wines from La Mancha Region — Determination of Key Odorants / Ed. by Dr. Bekir Salih. — In Tech Europe, Rijeka, Croatia. P. 147 – 170.
12. State Standard GOST R ISO 8588–2008. Organolepticheskii analiz. Metodologiya. Ispytaniya “A” — “Ne A” [Sensory analysis. Methodology. Trials “A” — “Not A”]. — Moscow: Standartinform, 2009. — 12 p. [in Russian].
13. State Standard GOST R 53161–2008. Organolepticheskii analiz. Metodologiya. Metod parnogo sravneniya [Sensory analysis. Methodology. The method of paired comparison]. — Moscow: Standartinform, 2009. — 20 p. [in Russian].
14. State Standard GOST R 53159–2008. Organolepticheskii analiz. Metodologiya. Metod treugol'nika [Sensory analysis. Methodology. The method of triangle]. — Moscow: Standartinform, 2009. — 16 p. [in Russian].