

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ЗАДАЧАХ ВИРТУАЛИЗАЦИИ

© 2020 г. А. А. Моисеев*

НПП “Технос-РМ”, Мытищи, 141002, Россия

*e-mail: slow.coach@yandex.ru

Поступила в редакцию 23.04.2019 г.

После доработки 31.08.2019 г.

Принята к публикации 17.09.2019 г.

Важной составляющей виртуализационного инструментария является статистическая составляющая для решения следующих задач: идентификации статистических моделей по результатам квалификационных испытаний, параметризации критериальных моделей для проведения прогнозирования в рамках виртуальной квалификации, синтеза регрессионных моделей, статистического анализа их адекватности и последовательной оптимизации в ходе факторного планирования. При этом линейность регрессионной модели позволяет осуществить контроль значимости факторов и полноты их системы с использованием традиционных статистических методов, а также оптимизацию параметров с помощью методов линейного программирования.

В ходе идентификации измерений характеристик нефтепродуктов показана ненормальность распределений их погрешности. Для построения рациональных алгоритмов обработки этих измерений был синтезирован ряд альтернативных аппроксимаций. Для аппроксимации бимодального распределения плотности использована сумма различающихся нормальных, а для аппроксимации одномодальных распределений прочих характеристик – аппроксимация по Вейбуллу. Параметризация построенных распределений осуществлялась методом моментов, что обеспечило удовлетворительность построенных аппроксимаций. Для оценки воспроизводимости результатов виртуальной квалификации были построены аналитические распределения статистики Кохрейна на базе F -распределения Фишера и статистики Граббса на базе T -распределения Стьюдента. Показано, что предварительная нормализация позволяет избежать необоснованного заключения о неоднородности исследуемых выборок и тем самым снижает вероятность ошибки первого рода при принятии решения о воспроизводимости результатов квалификационного эксперимента. Параметризацию критериальной модели по результатам квалификационных испытаний предложено проводить по результатам прогнозирования эффекта в экстремальных условиях, что позволяет сформировать гарантированные квалификационные нормативы. Предварительный выбор этих условий предложено осуществлять методом последовательных приближений в рамках процедуры факторного анализа. Разработанная методика синтеза широкого класса критериев подобия базируется на интерпретации логарифма критерия подобия как линейной формы от логарифмов, трактуемых как независимые факторы. Оценка коэффициентов указанной формы – нагрузок факторов – осуществляется с использованием факторного анализа. Разработан также интуитивный метод ранжирования факторов в рамках линейно-регрессионной модели с использованием насыщенных планов и последующем отсеивании наименее нагруженных факторов. Его основным преимуществом является использование при анализе только нефиктивных факторов, а основным же недостатком – отсутствие регулярной методики установления порогов значимости.

Ключевые слова: инструментарий, идентификация, параметризация, критериальная модель, прогнозирование, регрессионная модель, факторное планирование, линейное программирование, бимодальное распределение, аппроксимация, статистика Кохрейна, статистика Граббса, квалификационные нормативы, критерий подобия, число подобия, фиктивный фактор, значимость, ранжирование

DOI: 10.1134/S2304487X19060063

Основой для проведения виртуализации служит виртуализационный инструментарий, включающий семейства базовых функциональных средств и базовых математических моделей [1]. Функциональные средства обычно используются для синтеза или симуляции программного обес-

печения различного назначения, включая обработку результатов квалификационных испытаний. Базовые модели используются для тестирования программного обеспечения, синтеза идентификационных моделей для модельно-ориентированных систем управления, а также алго-

Виртуализационный инструментарий



Рис. 1. Виртуализационный инструментарий.

ритмов обработки результатов квалификационных испытаний.

Важной составляющей виртуализационного инструментария, приведенного на рис. 1, является статистическая составляющая для решения следующих задач: идентификации статистических моделей по результатам квалификационных испытаний, параметризации критериальных моделей для проведения прогнозирования в рамках виртуальной квалификации, синтеза регрессионных моделей, статистического анализа их адекватности и последовательной оптимизации в ходе факторного планирования. Регрессионная модель в составе инструментария, калибруемая по результатам квалификационных испытаний, используется для планирования квалификационных экспериментов. Как правило, это планирование базируется на использовании насыщенных планов и предоставляет следующие возможности:

- исследовать значимость и полноту системы введенных факторов;

- осуществить последовательную оптимизацию параметров, соответствующих значимым факторам, в ходе квалификационных испытаний;

- осуществить прогнозирование результатов квалификационных испытаний с использованием откалиброванной регрессионной модели.

При этом линейность используемой регрессионной модели позволяет осуществить контроль значимости факторов и полноты их системы с использованием традиционных статистических методов, а оптимизацию параметров с помощью методов линейного программирования.

Статистический анализ адекватности регрессионной модели предполагает использование ряда статистических критериев. В работе [2] проведено усовершенствование методов Кохрейна и Граббса, обеспечивающее их робастность и базирующееся на предварительной нормализации исследуемой выборки с помощью соответствующего преобразования Смирнова. Показано, что разработанная процедура обеспечения робастности снижает вероятность ошибки первого рода при

принятии решения о воспроизводимости результатов квалификационного эксперимента.

Распределение статистики Кохрейна на базе F -распределения Фишера и статистики Граббса на базе T -распределения Стюдента построены из следующих соображений. По нормализованным выборкам отношений Стюдента рассчитываются дисперсии, которые затем анализируются по критерию Кохрейна [2]. Решающая статистика Кохрейна представляет собой отношение максимальной частной дисперсии $\max_{i=1..l} S_i^2$ и суммы част-

ных дисперсий $S_1^2 + \dots + S_l^2$ отношений Стюдента. Здесь l — число выборок измерений, по которым рассчитываются частные дисперсии для дробей Стюдента. Каждая из них представляет собой несмещенную выборочную оценку вида:

$$S_i^2 = \frac{S_{i1}^2 + \dots + S_{il}^2}{l-1},$$

$$S_{ij}^2 = \left(\xi_{ij} - \frac{1}{p} \sum_{k=1}^p \xi_{ik} \right)^2,$$

которая может быть преобразована следующим образом:

$$C = \frac{\max_{i=1..l} S_i^2}{S_1^2 + \dots + S_l^2} = \max_{i=1..l} C_i,$$

$$C_i = \frac{S_i^2}{S_1^2 + \dots + S_l^2}.$$

В предположении о нормальности распределения числитель в C_i распределен по χ^2 со степенью свободы $np - 1$, а знаменатель представляет собой сумму l величин с таким распределением, и следовательно, распределен по χ^2 со степенью свободы $l(np - 1)$. Отсюда следует, что статистика

$$F_i = lC_i = \frac{\frac{S_i^2}{np-1}}{\frac{S_1^2 + \dots + S_l^2}{l(np-1)}} \quad \text{имеет } F\text{-распределение}$$

Фишера со степенями свободы $np - 1$ и $l(np - 1)$. Статистика Кохрейна принимает, таким образом,

$$\text{вид } C = \frac{F_M}{l}, \text{ где } F_M = \max_{i=1..l} F_i.$$

Аналогичным образом строится критерий Граббса для нормализованных отношений Стюдента. Эти отношения образуют вариационный ряд $\xi_1, \dots, \xi_p$, решающими статистиками в котором являются минимальное и максимальное отношения Стюдента: $G_{\min} = \xi_1$, $G_{\max} = \xi_p$. В случае нормальности выборки отношения Стюдента распределены по Стюденту, а порядковые статистики $G_{\min}$ и $G_{\max}$ имеют распределения вида $T_{n-1}^p(-G_{\min})$ и $T_{n-1}^p(G_{\max})$, где T_{n-1} — распределение

Стюдента с $n - 1$ степенью свободы. Решение об отсутствии воспроизводимости принимается, если выполняется условие $T_{n-1}^p(-G_{\min}) > P$ или $T_{n-1}^p(G_{\max}) > P$, где P — заданная доверительная вероятность. В противном случае считается, что измерения воспроизводимы.

Эквивалентным критерию Граббса является критерий Хокинса, решающая статистика которого H связана со статистиками $G_{\max}$, $G_{\min}$ соотношением:

$$H = \max(|G_{\max}|, |G_{\min}|).$$

С учетом того, что H является порядковой статистикой, ее распределение в предположении о нормальности генеральной совокупности измерений имеет вид $(2T_{n-1}(H) - 1)^p$. Соответствующее решающее правило при этом таково: принимается решение об отсутствии воспроизводимости измерений, если $(2T_{n-1}(H) - 1)^p > P$, и об их воспроизводимости в противном случае.

Важной составляющей межлабораторных измерений характеристик нефтепродуктов является идентификация распределения их погрешности, определяющего методику дальнейшей обработки измерений. В статье [3] проведен анализ распределения этих погрешностей, включая измерения плотности, вязкости при температурах 20°C и 100°C, а также температур отгона и вспышки. Анализ погрешностей измерений этих характеристик, проведенный методом χ^2 , показал, что используемое обычно предположение о нормальности этого распределения представляется неосновательным. В этих условиях возникает необходимость провести идентификацию указанного закона распределения и на этой основе разработать алгоритм рациональной обработки измерений. Решению первой части этой задачи посвящена работа [4], в которой исследуются возможные аппроксимации распределения погрешности измерений.

Рассмотрим некоторые методы построения аппроксимирующих распределений и их параметризации. Исходным пунктом построения аппроксимирующего распределения погрешности измерения плотности являлось построение частотной гистограммы указанной погрешности, отображенной на рис. 2. Она явно указывает на бимодальный характер распределения погрешности соответствующих измерений. Для аппроксимации этого распределения предложено использовать распределение вида:

$$f(x) = \frac{\exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right)}{2\sigma_1\sqrt{2\pi}} + \frac{\exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2}\right)}{2\sigma_2\sqrt{2\pi}}. \quad (1)$$

Соответствующее интегральное распределение представляется при этом в форме:

$$F = \frac{1}{2} \Phi \left(\frac{x - m_1}{\sigma_1} \right) + \frac{1}{2} \Phi \left(\frac{x - m_2}{\sigma_2} \right). \quad (2)$$

Мы видим, что аппроксимирующее распределение включает следующие параметры:

– m_1, m_2 – математические ожидания составляющих распределения;

– σ_1, σ_2 – СКО составляющих распределения.

Для оценки этих параметров предлагается использовать метод моментов. Он позволяет получить для них следующую систему алгебраических уравнений:

$$\begin{cases} \frac{m_1 + m_2}{2} = m; \\ \sigma_1^2 = \sigma^2 - \frac{(m_2 - m_1)^2}{2} - \frac{A}{3 \left(\frac{m_2 - m_1}{2} \right)}; \\ 3 \left(\frac{m_2 - m_1}{2} \right) \frac{\sigma_2^2 - \sigma_1^2}{2} = A; \\ 3 \frac{\sigma_1^4 + \sigma_2^4}{2} + \left(\frac{m_2 - m_1}{2} \right)^4 + 6 \left(\frac{m_2 - m_1}{2} \right)^2 \frac{\sigma_1^2 + \sigma_2^2}{2} = E. \end{cases} \quad (3)$$

Здесь A и E – выборочные асимметрия и эксцесс для выборки погрешностей измерений. Решение построенной системы уравнений осуществлялось методом перебора по параметру m_1 в диапазоне $(0, m)$. Значение m_2 находилось для данного m_1 из первого уравнения системы (3), σ_1^2 – из второго по заданному m_1 и рассчитанному m_2 , а σ_2^2 – из четвертого уравнения, которое при заданном m_1 и рассчитанных m_2, σ_1^2 представляет собой квадратное уравнение относительно σ_2^2 . В ходе варьирования параметр m_1 подбирался так, что бы обеспечить минимальные отклонения вычисленных параметров от выборочных. Результат проведенного таким образом расчета явились следующие значения параметров:

$$\begin{aligned} m_1 &= -0.353; \\ m_2 &= 0.314; \\ \sigma_1 &= 1.384; \\ \sigma_2 &= 0.519. \end{aligned}$$

Подставляя эти значения в (1) и (2), получаем аппроксимирующие распределения погрешности измерений плотности. Результат сравнения интегрального аппроксимирующего распределения с выборочным приведен на рис. 2 и указывает на качественно удовлетворительную аппроксимацию распределения указанной погрешности.

Частотные гистограммы для стандартизированной погрешности измерений прочих характеристик нефтепродуктов – вязкости при температурах 20°C и 100°C, а также температур отгона и вспышки – приведены на рис. 3 и указывают на унимодальный характер распределений погрешности их измерений. В качестве возможных аппроксимаций этих распределений в работе [4] рассматривались наряду с нормальной аппроксимацией по Пирсону и Вейбуллу, а также логнормальная. Параметризация этих распределений также осуществлялась методом моментов. Проведенный сравнительный анализ построенных аппроксимаций унимодальных распределений погрешности показывает, что наилучшей по точности является аппроксимация по Вейбуллу. Сравнение полученных распределений погрешности с бимодальной аппроксимацией для погрешности измерения плотности и вейбулловской аппроксимацией для погрешности измерения прочих характеристик указывает на удовлетворительность используемых аппроксимаций.

Модифицированный алгоритм контроля воспроизводимости межлабораторных измерений базируется на выравнивании последних путем перехода к дробям Стьюдента, что позволяет осуществлять контроль воспроизводимости по измерениям изменяющихся величин [5]. Однако в случае нормальной выборки измерений распределение дробей Стьюдента оказывается стандартным и не может быть использовано для контроля воспроизводимости. Поэтому отношение Стьюдента целесообразно определить в виде $\xi = \frac{x - m}{s(m)}$,

где $s(m)$ – аппроксимация зависимости СКО от математического ожидания. При этом трансформация измерений в отношения Стьюдента предусматривает предварительный расчет математических ожиданий и СКО, соответствующие разным уровням. По ним строится функция $s(m)$ линейной регрессии, которая совместно с расчетными математическими ожиданиями используется для построения дробей (отношений) Стьюдента.

Дальнейшие статистические выводы опираются на анализ построенных дробей Стьюдента, которые анализируются по критерию Граббса. В случае невоспроизводимости измерений по критерию Граббса серия, соответствующая экстремальной статистике, бракуется. Забракованная серия удаляется из выборки измерений, а решающие статистики пересчитываются для оставшихся уровней в предположении, что аппроксимация $s(m)$ остается неизменной. Это можно интерпретировать как замену забракованной серии такой, которая не изменяет аппроксимацию, т.е. мы пренебрегаем влиянием аномальной серии на последнюю. Если и в усеченной выборке оказывается возможным выделить аномальную серию,

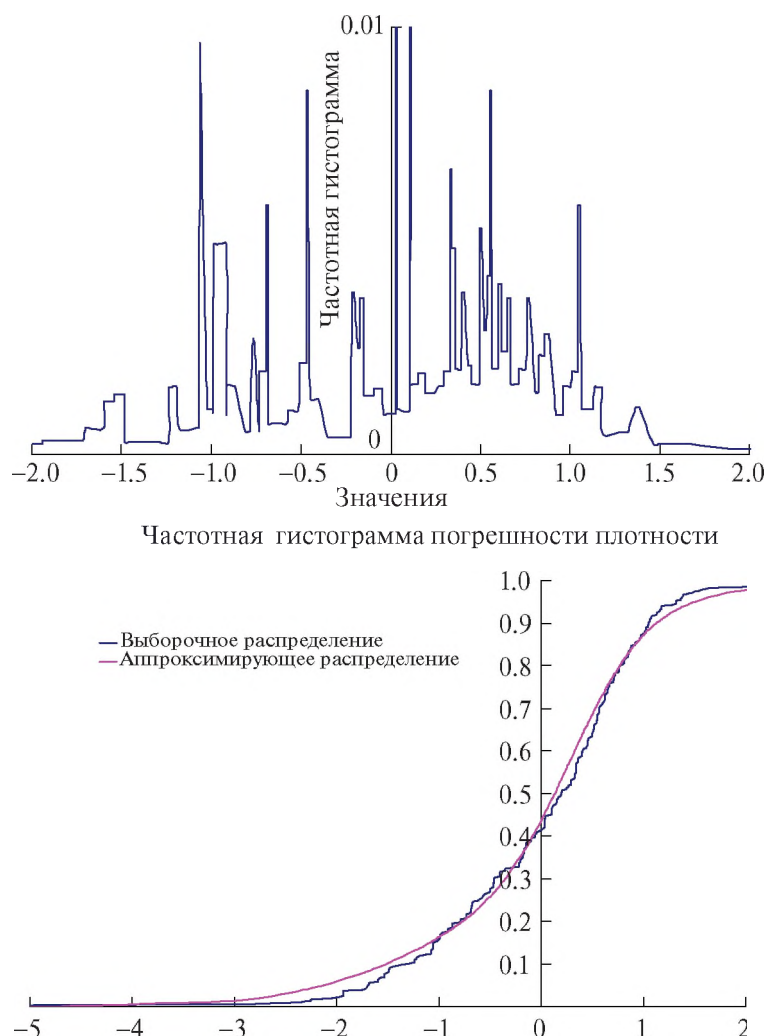


Рис. 2. Аппроксимация распределения погрешности измерения плотности.

предположение о слабом влиянии отброшенной серии на аппроксимацию считается несостоятельным и необходимо осуществить перерасчет аппроксимации без учета забракованной серии.

По сформированным таким образом выборкам отношений Стьюдента рассчитываются дисперсии, которые затем анализируются по критерию Кохрейна. Она является порядковой и имеет распределение $F^l(F_M, np - 1, l(np - 1)) = F^l(IC, np - 1, l(np - 1))$, где F – распределение Фишера. Условие неоднородности в этих условиях имеет вид $F^l(IC, np - 1, l(np - 1)) > P$, где P – доверительная вероятность решения о невоспроизводимости. При невыполнении этого условия принимается решение об однородности выборки, и, следовательно, воспроизводимости. Таким образом, модифицированный алгоритм контроля воспроизводимости применяется в условиях возможного изменения измеряемых величин. В этом смысле

он отличается универсальностью от обычной методики классификации, ориентированной на неизменность измеряемых значений. Использование выравнивания измерений путем перехода к дробям Стьюдента позволяет применять эту методику без всяких ограничений.

В соответствии с [6] критериальная модель позволяет осуществить прогнозирование по результатам квалификационных испытаний, проводимых в заданных условиях. Возникает, однако, вопрос о выборе таких условий, которые обеспечивали бы гарантированный прогноз – например, его верхнюю границу. Отсюда вытекает задача предварительного выбора экстремальных условий квалификационных испытаний, для решения которой предлагается использовать метод факторного планирования эксперимента. Статистическая обработка этих измерений, отображенная на рис. 4, включает:

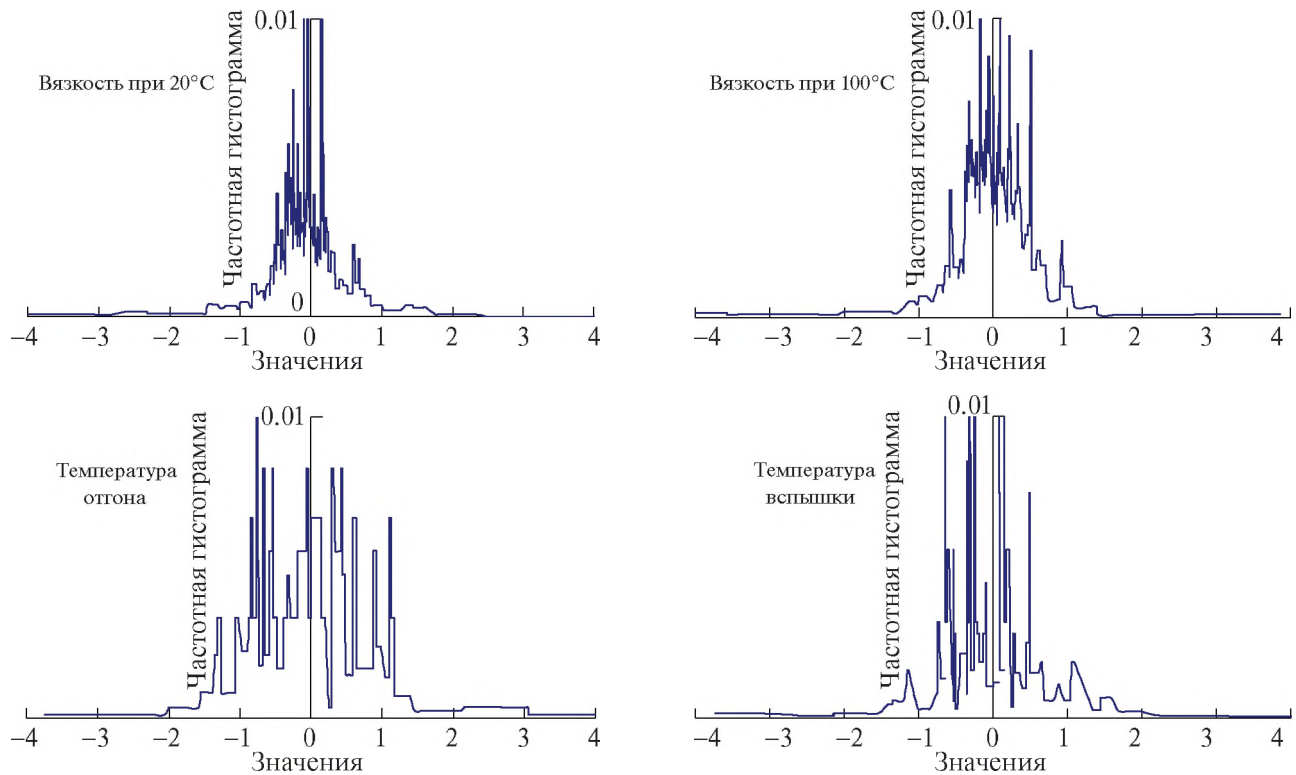


Рис. 3. Частотные гистограммы погрешности измерений.

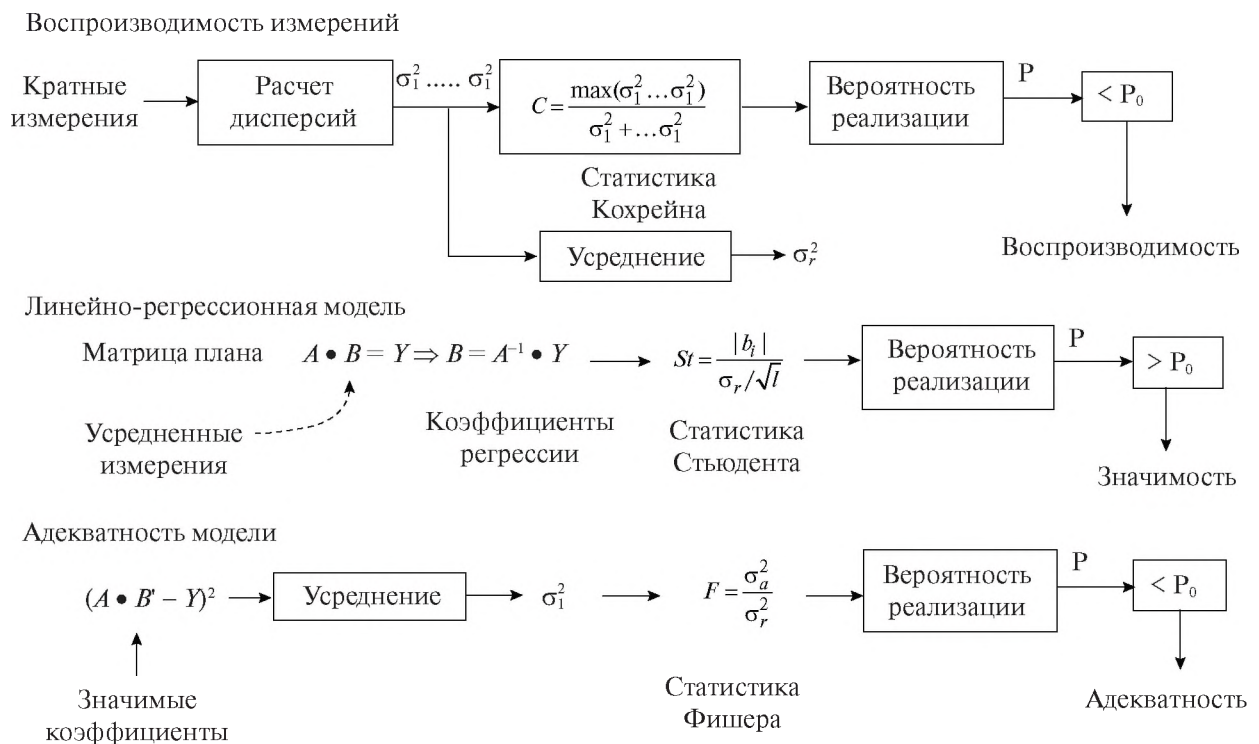


Рис. 4. Статистическая обработка измерений.

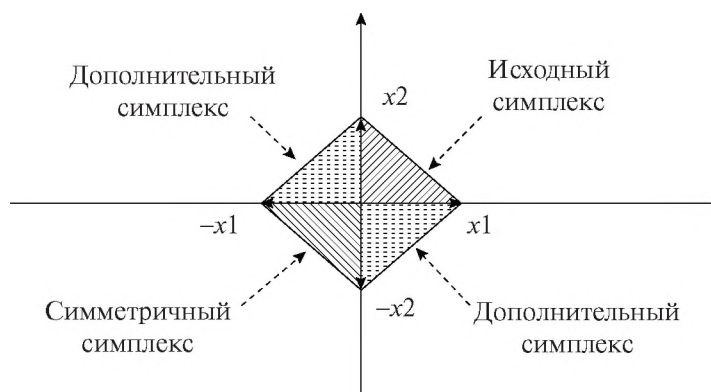


Рис. 5. Расширение факторной области.

- оценку воспроизводимости кратных измерений путем анализа их разброса;
- построение линейно-регрессионной модели и оценку значимости введенных факторов;
- анализа адекватности модели и оценку полноты системы введенных факторов.

Для оценки воспроизводимости измерений используется критерий Кохрейна. Он предусматривает расчет дисперсий кратных измерений, формирование по ним статистики Кохрейна и расчет вероятности ее реализации. Эта вероятность сравнивается с доверительной вероятностью невоспроизводимости измерений. Если рассчитанная вероятность оказывается ниже доверительной, принимается гипотеза о воспроизводимости измерений. Их дисперсии при этом усредняются, образуя дисперсию воспроизводимости. В противном случае серия измерений считается аномальной и должна быть повторена.

Расчет коэффициентов регрессии осуществляется из следующих соображений. Матрица плана для нефиктивных факторов является невырожденной матрицей системы линейных уравнений, правой частью которой является вектор усредненных измерений. Вектор коэффициентов, интерпретируемых как нагрузки факторов, представляет собой решение указанной системы.

Для оценки значимости факторов используются стьюдентовы отношения, решение по которым принимается по критерию Стьюдента. С этой целью рассчитываются статистики Стьюдента для коэффициентов регрессии и вероятности их реализации. Значимыми считаются коэффициенты, вероятности реализации статистики Стьюдента для которых превышают доверительную. Прочие коэффициенты регрессии считаются незначимыми и в модели должны быть обнулены.

Контроль адекватности линейно-регрессионной модели (полноты системы введенных факторов) осуществляется по критерию Фишера с использованием отношения дисперсии адекватно-

сти к дисперсии воспроизводимости. Для этого рассчитывается средний квадрат отклонения усредненных измерений от регрессионной модели, в которой сохраняются только значимые коэффициенты регрессии. Эта величина интерпретируется как дисперсия адекватности. По ней и дисперсии воспроизводимости строится статистика Фишера и рассчитывается вероятность ее реализации. Решение об адекватности принимается, если указанная вероятность ниже доверительной. В противном случае модель считается неадекватной, что, как правило, означает потерю существенных факторов при ее построении.

Двухуровневое факторное планирование может быть использовано для последовательной оптимизации параметров исследуемого объекта или квалификационного эксперимента. Поскольку локальные оптимумы при этом достигаются в вершинах многомерных симплексов, соответствующих строкам двухуровневого плана, поиск глобального экстремума предлагается осуществлять методом последовательных приближений. Последовательная оптимизация предусматривает использование точек достигнутых локальных экстремумов в качестве центров новых областей в параметрическом пространстве.

При использовании линейно-регрессионной модели поиск экстремума представляет собой задачу линейного программирования для замкнутой области факторных значений в виде многомерного симплекса. Локальный экстремум линейной регрессии достигается при этом в одной из вершин области и, следовательно, равен экстремальному усредненному измерению, соответствующему одной из строк плана эксперимента. Симплекс-планирование предусматривает включение в факторную область также симплекса, симметричного рассмотренному. Ему соответствует план эксперимента с инвертированной матрицей факторов, а общая область поиска экстремума согласно рис. 5 образуется исходным,

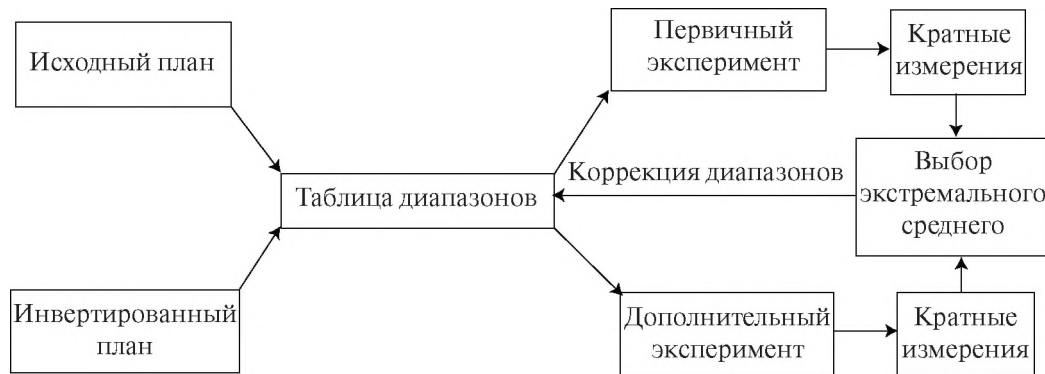


Рис. 6. Статистическая обработка измерений.

симметричным и дополнительными симплексами, примыкающими к началу координат.

При расширенном планировании, отображенном на рис. 6, для формирования таблиц параметров используются исходный и инвертированный планы. В соответствии с этими таблицами проводятся квалификационные испытания, в ходе которых осуществляются кратные измерения. Обработка этих измерений осуществляется в два этапа. В ходе первичной обработки осуществляется поиск экстремального усредненного измерения, параметры которого сопоставляются новому центру диапазонов. Этому центру соответствует скорректированный диапазон параметров, используемый для реализации следующего приближения процедуры поиска. Если достигнутый в следующем приближении экстремум “экстремальнее” предыдущего, процедура продолжается. В противном случае она останавливается и глобальному экстремуму сопоставляется значение предыдущего локального. Параметризация критериальной модели по результатам квалификационных испытаний в экстремальных условиях позволяет, например, провести прогнозирование фрикционного износа в этих условиях и сформировать на этой основе гарантированные нормы износа.

Интересным примером применения статистических методов в задачах виртуализации является факторное планирование в анализе критериев подобия [7]. Критериальные модели являются естественной базой для виртуализации квалификационных испытаний, поскольку последние базируются на физической модели исследуемого объекта или процесса в виде лабораторной установки. Эксперименты, проводимые на ней, позволяют осуществить параметризацию критериальной модели, чтобы затем использовать ее для пересчета лабораторных результатов в эксплуатационные характеристики, а также для обратного пересчета эксплуатационных условий в параметры квалификационных испытаний.

Важной задачей при обосновании инвариантности используемых критериев подобия является их представление в виде функций от стандартных чисел подобия, поскольку их инвариантность обеспечивает инвариантность построенного критерия. Как правило, эта функция является мультипликативно-степенной, т.е. представляется в виде:

$$K = K_0 N_1^{\beta_1} \dots N_l^{\beta_l}, \quad (4)$$

где $N_1, \dots, N_l$ — используемые числа подобия, а $\beta_1, \dots, \beta_l$ — их степени в составе критерия.

Логарифмирование критерия (3) трансформирует мультипликативно-степенную функцию в линейную комбинацию $\ln K_0 + \beta_1 \ln N_1 + \dots + \beta_l \ln N_l = \ln K$. Логарифмы чисел подобия интерпретируются как натуральные параметры, стандартным образом преобразуемые в двухуровневые факторы, а построенная выше линейная комбинация для натуральных параметров с априорно неопределенными коэффициентами трансформируется в линейную регрессию для этих факторов. Ее коэффициенты подлежат определению по результатам калибровочных экспериментов. Для решения этой задачи традиционно используется двухуровневое факторное планирование. Факторы, определяемые на основе натуральных параметров, являются нефиктивными, но могут быть дополнены фиктивными факторами в составе линейно-регрессионной модели.

Матрица двухуровневого планирования является матрицей коэффициентов в линейной системе уравнений для коэффициентов регрессии, включая фиктивные. Правой частью этой системы являются “измерения” — значения логарифма критерия подобия, соответствующие границам натуральных параметров в составе линейной комбинации. Анализ значимости рассчитываемых коэффициентов регрессии и адекватности линейно-регрессионной модели в целом осуществляются стандартными статистическими методами.

В качестве примера предложенного подхода рассмотрим анализ критерия фрикционного износа в смазывающей жидкости. Данный критерий имеет вид [8]:

$$V_w = \frac{\rho v \Delta V}{\mu N t}.$$

Здесь ρ — плотность изнашивающегося материала, v — скорость скольжения, ΔV — величина объемного износа, μ — коэффициент трения, N — нормальная нагрузка на трущуюся поверхность, а t — длительность износа. Физический смысл этого критерия — отношение импульса износного материала к импульсу силы трения.

Представим данный критерий в виде линейной функции чисел подобия:

$$\ln(V_w) = \beta_0 + \beta_1(\ln St) + \beta_2(\ln So) + \beta_3(\ln Re) + \beta_4(\ln De),$$

где e^{β_0} — коэффициент пропорциональности; $St = \frac{v t}{D}$ — число Струхала (гомохронности), D — максимальный диаметр полосы износа шара; $So = \frac{\eta v}{\mu \frac{N}{D^2} \Delta L}$ — число Зоммерфельда, η — динамическая вязкость жидкости, ΔL — глубина износа шара; $Re = \frac{\rho_f v \Delta L}{\eta}$ — число Рейнольдса, ρ_f — плотность жидкости; $De = Re \sqrt{\frac{d}{D}}$ — число Дина, d — минимальный диаметр полосы износа шара.

Для пересчета натуральных параметров в соответствующие факторы используем соотношения:

$$\begin{aligned} x_1 &= 2 \frac{\ln St - (\ln St)_{\min}}{(\ln St)_{\max} - (\ln St)_{\min}} - 1; \\ x_2 &= 2 \frac{\ln So - (\ln So)_{\min}}{(\ln So)_{\max} - (\ln So)_{\min}} - 1; \\ x_3 &= 2 \frac{\ln Re - (\ln Re)_{\min}}{(\ln Re)_{\max} - (\ln Re)_{\min}} - 1; \\ x_4 &= 2 \frac{\ln De - (\ln De)_{\min}}{(\ln De)_{\max} - (\ln De)_{\min}} - 1. \end{aligned} \quad (5)$$

Границы натуральных параметров, рассчитанные по результатам упомянутых выше экспериментов, приведены в [7]. При этом граничным значениям натуральных параметров соответствуют значения факторов -1 и 1 . Обратный пересчет факторов в натуральные параметры осуществляется в соответствии с соотношениями:

$$\ln St = (\ln St)_{\min} + ((\ln St)_{\max} - (\ln St)_{\min}) \frac{x_1 + 1}{2};$$

$$\ln So = (\ln So)_{\min} + ((\ln So)_{\max} - (\ln So)_{\min}) \frac{x_2 + 1}{2};$$

$$\ln Re = (\ln Re)_{\min} + ((\ln Re)_{\max} - (\ln Re)_{\min}) \frac{x_3 + 1}{2};$$

$$\ln De = (\ln De)_{\min} + ((\ln De)_{\max} - (\ln De)_{\min}) \frac{x_4 + 1}{2}.$$

Подставляя их в (4), получаем для регрессии критерия относительно нефиктивных факторов $\ln(V_w) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$, где:

$$\begin{aligned} b_0 &= \beta_0 + \frac{(\ln St)_{\max} + (\ln St)_{\min}}{2} + \\ &+ \frac{(\ln So)_{\max} + (\ln So)_{\min}}{2} + \\ &+ \frac{(\ln Re)_{\max} + (\ln Re)_{\min}}{2} + \\ &+ \frac{(\ln De)_{\max} + (\ln De)_{\min}}{2}; \\ b_1 &= \beta_1 \frac{(\ln St)_{\max} - (\ln St)_{\min}}{2}; \\ b_2 &= \beta_2 \frac{(\ln So)_{\max} - (\ln So)_{\min}}{2}; \\ b_3 &= \beta_3 \frac{(\ln Re)_{\max} - (\ln Re)_{\min}}{2}; \\ b_4 &= \beta_4 \frac{(\ln De)_{\max} - (\ln De)_{\min}}{2}. \end{aligned} \quad (6)$$

Построение регулярного плана для анализа влияния введенных факторов осуществляется на основе дробной реплики полного факторного эксперимента. При этом число m факторов в составе этого эксперимента должно удовлетворять условию $2^{m-1} < l + 1 \leq N = 2^m$, где l — определенное выше число нефиктивных факторов, а N — общее число факторов, равное ближайшей мажорирующей степени двойки. Учитывая это обстоятельство, необходимо ввести в соотношение для регрессии дополнительные фиктивные факторы x_5, x_6, x_7 . В результате оно приобретает вид:

$$\ln(V_w) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_6 + b_7 x_7.$$

Матрица планирования при этом соответствует дробной реплике 2^{7-4} полного факторного плана, которая приведена на рис. 7. Характерным свойством этой матрицы является центрированность ее столбцов, за исключением первого. Отсюда вытекает, что свободный член b_0 в выражении для регрессии равен среднему правых частей уравнений линейной системы. Исходя из этого, примем его в дальнейшем равным результату усреднения $\ln(V_w)$ по всем проведенным экспе-

риментам. Усеченная матрица плана относительно коэффициентов регрессии для оставшихся факторов выделена на рис. 7. При расчете правых частей усеченной линейной системы уравнений относительно этих коэффициентов будем исходить из следующих соображений. Значение $\ln(Iw)$ для данного эксперимента будем сопоставлять значению фактора $x_i = -1$, если в соответствии с (4) $x_i < 0$, и значению $x_i = 1$ в противном случае. Таким образом, каждому значению $\ln(Iw)$ сопоставляет определенный набор факторов x_1, x_2, x_3, x_4 . Значения $\ln(Iw)$ с одним и тем же набором факторов интерпретируются как повторяющиеся “измерения” для соответствующей строки плана. Для каждой такой группы “измерений” рассчитываются среднее и дисперсия.

Таким образом, для расчета коэффициентов регрессии $b_1, \dots, b_7$ используется усеченная матрица плана, а правая часть линейной системы представляет собой усредненные “измерения” за вычетом свободного члена. Вектор коэффициентов регрессии представляет собой решение этой линейной системы и рассчитывается путем умножения обращенной матрицы плана на определенный выше вектор правых частей. При этом следует учитывать, что факторы x_5, x_6, x_7 являются фиктивными и сумма квадратов соответствующих им коэффициентов суммируется с усредненной дисперсией, рассчитанной по повторяющимся “измерениям”, давая дисперсию воспроизводимости σ^2 . Относя ее к количеству нефиктивных факторов $l = 4$, мы получаем дисперсию оценки соответствующих коэффициентов регрессии и ее среднеквадратичное отклонение $s_b = \frac{\sigma}{\sqrt{l}}$.

Для оценки значимости коэффициентов регрессии при нефиктивных факторах $b_1, \dots, b_4$ используются статистики $\frac{|b_1|}{s_b}, \dots, \frac{|b_4|}{s_b}$, распределенные по Стюденту с числом степеней свободы l . Вероятности реализации этих статистик составляют $T_l\left(\frac{|b_1|}{s_b}\right), \dots, T_l\left(\frac{|b_4|}{s_b}\right)$, где T_l – распределение Стюдента с l степенями свободы. Решающее правило для подтверждения значимости i -го коэффициента имеет вид $T_l\left(\frac{|b_i|}{s_b}\right) > P = 0.9$, где P – доверительная вероятность принятия решения о значимости. Коэффициенты регрессии, для которых эти условия не выполняются, считаются незначимыми и приравниваются нулю. Результаты расчетов, отображенные на рис. 4, показывают, что значимыми, т.е. ненулевыми, являются только коэффициенты b_2 и b_4 , соответствующие числам подобия Зоммерфельда и Дина. Прочие коэффици-

циенты являются незначимыми и поэтому зависимость от критериев Струхала и Рейнольдса игнорируется.

Для контроля адекватности полученной регрессии рассчитывается рассогласование между усредненными “измерениями” и значениями регрессии в каждой строке плана. Для расчета этих значений строка плана умножается на вектор значимых коэффициентов регрессии, отображенный на рис. 7. Решение об адекватности линейно-регрессионной модели принимается по критерию Кохрейна. Решающей при этом является статистика Кохрейна C – отношение максимального квадрата отклонения к сумме квадратов отклонений. Вероятность реализации этой статистики составляет $F^N(C, 1, N)$, где F – распределение Фишера. Решающее правило при этом имеет следующий вид: принимается решение об адекватности, если вероятность реализации ниже доверительного значения 0.9, и об отсутствии адекватности в противном случае. Результаты расчета статистики C и вероятности ее реализации также приведены на рис. 7. Поскольку полученная вероятность ниже доверительного значения, принимается гипотеза об адекватности построенной линейно-регрессионной модели.

Для формирования окончательного вида критерия подобия необходимо предварительно преобразовать значимые коэффициенты b_2 и b_4 в коэффициенты β_2 и β_4 , входящие в (3). Это осуществляется путем обращения соответствующих выражений в (5):

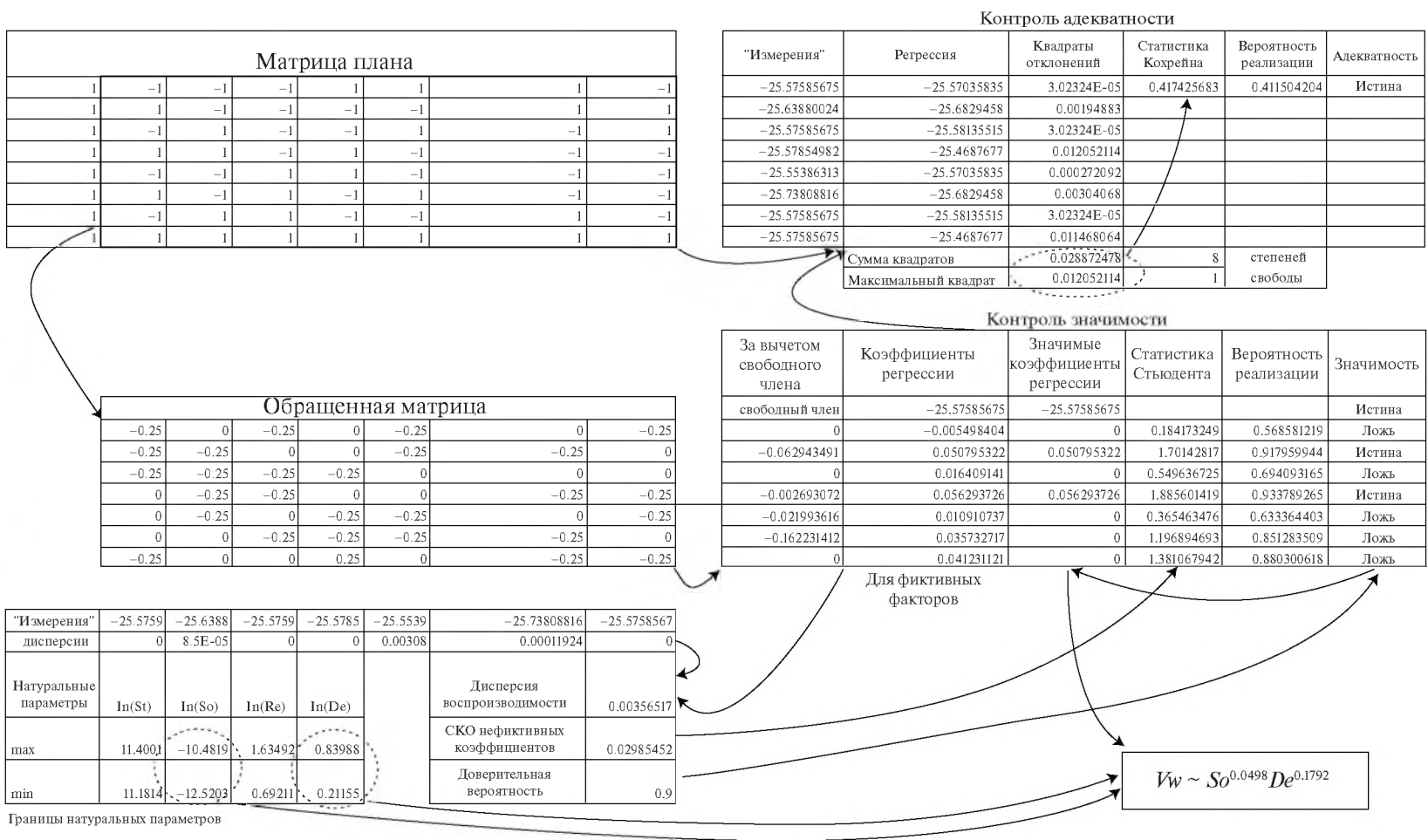
$$Vw \sim So^{\beta_2} De^{\beta_4}$$

$$\beta_2 = \frac{2b_2}{(\ln So)_{\max} - (\ln So)_{\min}} = 0.0498;$$

$$\beta_4 = \frac{2b_4}{(\ln De)_{\max} - (\ln De)_{\min}} = 0.1792.$$

Проведенное рассмотрение показывает, что факторное планирование может быть использовано для обоснования инвариантности широкого класса вводимых критериев подобия, а также для доказательства адекватности их представления в зависимости от стандартных чисел подобия. Преимуществом предложенного подхода является то, что он может быть использован практически без ограничений общности и базируется на хорошо развитых методах факторного и статистического анализа.

Обычная методика оценки прецизионности соответствует стандартному подходу к обработке комплексного эксперимента. Он отличается возможным различием величин, измеряемых в последовательных сериях. Основной в этих условиях является задача исследования влияния учитываемых факторов на измеряемую величину.



Исследование воспроизводимости является в этих условиях вспомогательным и предназначено для отделения влияния погрешности измерений от влияния учитываемых факторов. Иная ситуация возникает при анализе прецизионности межлабораторных испытаний. Исследование влияния факторов при этом становится вторичным, а главным становится исследование межлабораторной воспроизводимости [9]. Задачей межлабораторных сравнительных испытаний является анализ компетентности лабораторий, а также их выбраковка по признаку некомпетентности. Ниже разрабатывается статистический алгоритм анализа компетентности лабораторий, участвующих в указанных испытаниях. Алгоритм базируется на предварительной стандартизации проводимых измерений путем преобразования их в соответствующие отношения Стьюдента. Сформулированы условия некомпетентности лаборатории в случае аномально большой ошибки измерений и в случае нескольких умеренных ошибок.

В основе данного анализа лежит подход, предусматривающий сравнительные измерения разнотипных стандартных образцов веществ и материалов, а также их рациональную обработку с использованием методов дисперсионного анализа. Целью последнего является разложение суммарной дисперсии на две величины: дисперсию, обусловленную техникой эксперимента, и дисперсию, вызванную действием исследуемого фактора. В случае многофакторного эксперимента определяются дисперсии, соответствующие различным факторам и их взаимодействиям, а также оценивается статистическая значимость этих величин с учетом ошибки воспроизводимости.

Компетентность лаборатории определяется воспроизводимостью результатов сравнительных испытаний, которая обеспечивается в свою очередь статистической однородностью выборки проведенных измерений. Традиционными методами анализа этой однородности являются сравнительный анализ серийных дисперсий исследуемой выборки или серийных стьюдентовых отношений. Довольно часто для решения этой задачи применяются следующие параметрические критерии: Кохрейна, базирующийся на сравнении максимума серийных дисперсий с их суммой, и Граббса (или Хокинса), базирующийся на анализе аномальности экстремальных значений в серийных выборках стьюдентовых отношений для измерений.

Предварительным этапом обработки является стандартизация производимых измерений. Для данного стандартного образца она осуществляется путем перехода от измерений характеристики этого образца к дробям Стьюдента, построенным в соответствии с этими измерениями. Последние формируются в соответствии с соотношениями:

$$\xi_i = \frac{x_i - m}{s};$$

$$m = \frac{\sum_{k=1}^L x_k}{L};$$

$$s = \sqrt{\frac{\sum_{k=1}^L (x_k - m)^2}{L - 1}},$$

где x_i — межлабораторные измерения стандартного образца; m , s — величины среднего и СКО измерений; L — количество лабораторий-участниц.

В этих условиях для решения задачи обработки стандартизированных измерений естественно использовать критерий Хокинса с выборочными средними и СКО, рассчитываемыми по серийным выборкам. Решающими статистиками в этом случае являются $Y = \max_{i=1...n} \xi_i$ и $y = \min_{i=1...n} \xi_i$, а условия аномальности крайних измерений в серии принимают вид:

$$(2T_{L-1}(H) - 1)^L > P, \quad (7)$$

$$H = \max(|Y|, |y|).$$

Это условие можно интерпретировать как первое условие некомпетентности: лаборатория, в которой проведено соответствующее ξ измерение, считается некомпетентной.

Еще одно условие некомпетентности связано с подсчетом неудовлетворительных измерений из всех проведенных в данной лаборатории. Удовлетворительными считаются измерения, близкие к среднему m , а неудовлетворительными — сильно отклоняющиеся от него. Поскольку эти измерения относятся к различным образцам, они также предварительно стандартизируются путем преобразования в отношения Стьюдента. При переходе к ним удовлетворительными считаются значения дробей, расположенные в окрестности 0, а неудовлетворительными — лежащие вне этой окрестности.

Зададимся вероятностью p неудовлетворительности измерения. Учитывая, что дроби Стьюдента имеют T -распределение с $L - 1$ степенью свободы, находим для порога H принятия решения об указанной неудовлетворительности:

$$2T_{L-1}(H) - 1 = 1 - p. \quad (8)$$

Удовлетворительным считается значение дроби Стьюдента меньше H , а неудовлетворительным — большее H . Распределение вероятностей количества неудовлетворительных измерений n из N проведенных в данной лаборатории является биномиальным с параметрами p и N . Задаваясь доверительной вероятностью P_{II} выбраковки по числу неудовлетворительных измерений, полу-

Выборкам, включающим три фактора из пяти, соответствует коэффициент при i^3 , т.е. набор планов вида $x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2x_5 + x_1x_3x_5 + x_1x_4x_5 + x_2x_3x_5 + x_2x_4x_5$, соответствующих репликам 2^{5-3} . Для каждого из выбранных по такому принципу планов проводится серия троекратно повторяющихся экспериментов. Коэффициенты регрессии для каждого из этих планов рассчитываются однозначно в соответствии с (8), где $y_1, \dots, y_N$ интерпретируются как медианы повторяющихся измерений. В свою очередь, по найденным коэффициентам рассчитываются нагрузки для факторов, охватываемых данной серией. Поскольку каждый фактор охватывается $P - 1$ серией экспериментов, для него получается то же число относительных нагрузок $L_{i1}, \dots, L_{iP-1}$, рассчитываемых с учетом изменения диапазона измерений от серии к серии. Характеристиками значимости факторов считаются максимальные нагрузки из полученных $L_{imax} = \max(L_{i1}, \dots, L_{iP})$, $i = 1, \dots, k$, по которым и осуществляется ранжирование.

Относительные максимальные нагрузки $L_{1max}, \dots, L_{kmax}$ можно также использовать для рационального отсеивания незначимых факторов. Предположим без ограничения общности, что ряд $L_{1max}, \dots, L_{kmax}$ представляет собой результат ранжирования нагрузок по убыванию, и образуем эквивалентный ряд $\frac{L_{1max}}{L}, \dots, \frac{L_{kmax}}{L}$, где $L = L_{1max}$.

Выставим порог по значимости, например, $h = 0.05$, и все факторы, удовлетворяющие условию $\frac{L_{imax}}{L} < h$, будем считать незначимыми. Этот интуитивный метод отсеивания более прост в сравнении с традиционным статистическим отсеиванием. Его основным преимуществом, наряду с простотой реализации, является использование при анализе только нефиктивных факторов. Основным же недостатком, наряду с интуитивностью, является отсутствие регулярной методики выбора порога значимости, который приходится выбирать интуитивно.

Проведенное рассмотрение показывает, что статистическая составляющая виртуального инструментария охватывает широкий круг задач обработки результатов квалификации. Используемые при этом процедуры приведены к аналитическому виду и обеспечивают достаточную строгость анализа. Вместе с тем, они базируются на стандартных подходах и сравнительно просты в обращении.

ВЫВОДЫ

1. Важной составляющей виртуализационного инструментария является статистическая составляющая для решения следующих задач: иденти-

фикации статистических моделей по результатам квалификационных испытаний, параметризации критериальных моделей для проведения прогнозирования в рамках виртуальной квалификации, синтеза регрессионных моделей, статистического анализа их адекватности и последовательной оптимизации в ходе факторного планирования. При этом линейность регрессионной модели позволяет осуществить контроль значимости факторов и полноты их системы с использованием традиционных статистических методов, а также оптимизацию параметров с помощью методов линейного программирования.

2. Идентификация распределений измерений различных параметров нефтепродуктов указывает на их ненормальность. В качестве альтернативной аппроксимации бимодального распределения плотности было предложено использовать многопараметрическое распределение, представляющее собой сумму различающихся нормальных. Одномодальные распределения прочих характеристик аппроксимируются по Вейбуллу. Параметризация построенных распределений осуществлялась методом моментов, что обеспечивает удовлетворительность используемых аппроксимаций.

3. Для оценки воспроизводимости результатов квалификационных экспериментов были построены аналитические распределения статистики Кохрейна на базе F -распределения Фишера и статистики Граббса на базе T -распределения Стьюдента. Показано, что предварительная нормализация позволяет избежать необоснованного заключения о неоднородности исследуемых выборок. Тем самым снижается вероятность ошибки первого рода при принятии решения о воспроизводимости результатов квалификационного эксперимента.

4. Параметризацию критериальной модели по результатам квалификационных испытаний предложено проводить по результатам прогнозирования эффекта в экстремальных условиях, что позволяет сформировать гарантированные квалификационные нормативы. Предварительный выбор экстремальных условий испытаний при квалификационном нормировании предложено осуществлять методом последовательных приближений с использованием факторного планирования.

5. Предложен подход для обоснования инвариантности широкого класса критериев подобия. Подход базируется на интерпретации логарифма критерия как линейной формы от логарифмов чисел подобия, трактуемых как независимые факторы. Оценка коэффициентов указанной формы — нагрузок факторов — осуществляется с использованием факторного анализа и включает также доказательство адекватности указанного представ-

ления. Преимуществом предложенного подхода является возможность его применения без ограничений общности.

6. Разработан интуитивный метод ранжирования факторов в линейно-регрессионной модели с использованием насыщенных планов и последующем отсеивании наименее нагруженных факторов. Его основным преимуществом, наряду с простотой реализации, является использование при анализе только нефиктивных факторов, а основным недостатком — отсутствие регулярной методики установления порогов значимости, который приходится выбирать интуитивно.

СПИСОК ЛИТЕРАТУРЫ

1. Моисеев А.А. Виртуализационный инструментарий и его применение // Промышленные АСУ и контроллеры. 2016. № 11. С. 15.
2. Моисеев А.А. Модифицированные критерии статистической однородности // Промышленные АСУ и контроллеры. 2015. № 11. С. 20.
3. Шаталов К.В., Моисеев А.А. Проверка нормальности распределения погрешностей измерения при оценке качества нефтепродуктов / Труды 25 ГОСНИИ МО РФ, Вып. 57. М., “Перо”, 2016. С. 360.
4. Моисеев А.А. Аппроксимация распределения погрешности измерений характеристик нефтепродуктов // Промышленные АСУ и контроллеры. 2015. № 9. С. 30.
5. Моисеев А.А. Модифицированный алгоритм контроля воспроизводимости межлабораторных испытаний. 2016. № 9. С. 16.
6. Моисеев А.А. Критериальное моделирование в формировании квалификационных нормативов / Труды XI международной научно-технической конференции “Трибология — машиностроению”, М., 2016. С. 161.
7. Моисеев А.А. Факторное планирование в анализе критериев подобия // Вестник НИЯУ МИФИ. 2016. Т. 5. № 4. С. 303.
8. Моисеев А.А. Критериальная модель квалификационных испытаний на износ // Инженерная физика. 2015. № 12. С. 30.
9. Моисеев А.А. Статистический анализ компетентности при межлабораторных испытаниях // Моделирование, оптимизация и информационные технологии. 2017. № 1 (16), (<http://moit.vivt.ru/>).
10. Моисеев А.А. Интуитивный метод ранжирования факторов по значимости // Моделирование, оптимизация и информационные технологии. 2017. № 1 (16). (<http://moit.vivt.ru/>).

Vestnik Natsional'nogo issledovatel'skogo yadernogo universiteta “MIFI”, 2020, vol. 9, no. 1, pp. 66–81

Statistical Methods in Virtualization

A. A. Moiseev[#]

NPP Technos-RM (Technos-RM Research and Production Enterprise), Mytischy, 141002, Russia

[#]*e-mail: slow.coach@yandex.ru*

Received April 23, 2019; revised August 31, 2019; accepted September 17, 2019

Abstract—Essential component of virtual tool is statistical one used for statistical models identification on qualification tests results, criteria models parameterization for forecasting at qualification tests, regression models synthesis, their optimization and adequateness analysis at factor planning. Regression model linearity allows controlling factors significance using standard statistical methods and performing model optimization using linear programming. At oils characteristics identification their non-normality was shown. To form rational processing algorithms some alternative approximations were formed. Bimodal density distribution was approximated with sum of different normal ones. One — modal distributions of other characteristics were approximated with Weibull distributions. Parameterization of these distributions was performed using moment's method. For reproducibility estimation analytical Cochran distribution on the base of Fisher distribution and analytical Grabbs distribution on the base of Student distribution were developed. It was shown that preliminary normalization allows avoiding error decision about result reproducibility. Criteria model parameterization was proposed to perform on effect forecast at extremal conditions that allows to guaranty qualification normative. Preliminary choice of these conditions is realized by using step-by-step approach in the course of factor analysis. Developed method for similarity criteria synthesis is based on criterion logarithm presentation as linear form of some similarity numbers interpreted as independent factors. Coefficients of this form were estimated using factor analysis. Also the intuitive method of factors ranking in linear form using saturated plans and least significant factors rejection was developed. Its base advantage is fictive factors exclusion and base disadvantage — regular method absence for significance thresholds set.

Keywords: tool, identification, parameterization, criteria model, forecasting, regression model, factor planning, linear programming, bimodal distribution, approximation, Cochran statistics, Grabbs statistics, qualification normative, similarity criterion, similarity number, fictive factor, significance, ranking

DOI: 10.1134/S2304487X19060063

REFERENCES

1. *Moiseev A.* Virtualizacionnyj instrumentarij i ego primenenie [Virtualization's tool and its application] // Industrial Automatic Control Systems and Controllers, № 11, 2016, p. 16.
2. *Moiseev A.* Modifitsirovannye kriterii statisticheskoy neodnorodnosti [Modified criteria of statistical inhomogeneity] // Industrial Automatic Control Systems and Controllers, № 11, 2015, p. 20.
3. *Shatalov K., Moiseev A.* Proverka normalnosti raspredeleniya pogreshnostej izmereniya pri otsenke katshestva nefteproduktov [Normality control of measurements distribution at oils quality estimation] / Proceedings of State himmotology institute, issue 57, M., "Pero", 2016, p. 360.
4. *Moiseev A.* Approssimatsia raspredeleniya pogreshnosti izmerenij kharakteristik nefteproduktov [Error distributions approximations for measurements of oils characteristics] // Industrial Automatic Control Systems and Controllers, № 9, 2015, p. 30.
5. *Moiseev A.* Modifitsirovannyj algoritm kontrolya vosproizvodimosti mezhlaboratornykh ispytanij [Modified control algorithm of inter – laboratory tests reproducibility] // Industrial Automatic Control Systems and Controllers, № 9, 2016, p. 16.
6. *Moiseev A.* Kriterianoe modelirovaniye v formirovanii kvalifikatsionnykh normativov [Criteria modeling at qualification normative forming] / Proceedings of XI international scientific conference "Tribology to machine-building", M., 2016, p. 161.
7. *Moiseev A.* Faktornoe planirovanie v analize kriteriev podobiya [Factor's planning at similarity criteria analysis] // Vestnik Natsional'nogo issledovatel'skogo yadernogo universiteta "MIFI", 2016, vol. 5, no. 4, pp. 303.
8. *Moiseev A.* Kritikalnaya model kvalifikatsionnykh ispytanij na iznos [Criteria model of wear qualification tests] // Engineering physics, 2015, no. 12, p. 30.
9. *Moiseev A.* Statisticheskij analiz kopetentnosti pri mezhlaboratornykh ispytanoyakh [Statistical competence analysis at inter – laboratory tests] // Modeling, optimization and information technologies, 2017, № 1(16), (<http://moit.vivt.ru/>).
10. *Moiseev A.* Intuitivnyj metod ranzhirovaniya faktorov po znatshimosti [Intuitive method of factors ranking on significance] // Modeling, optimization and information technologies, 2017, № 1(16), (<http://moit.vivt.ru/>).