

Д. П. Халитов¹, Е. И. Молчанова¹

¹ *Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

ОЦЕНКА ВОЗМОЖНОСТИ СНИЖЕНИЯ КОЛИЧЕСТВА РУЧНЫХ ОПЕРАЦИЙ ПРИ ЭКСПЛУАТАЦИИ СИСТЕМЫ ФОРМИРОВАНИЯ ОТЧЕТНОСТИ В ИРКУТСКОМ ИВЦ

Аннотация. В данной статье осуществляется обзор эксплуатируемой Иркутским ИВЦ системы формирования отчетности, а также рассмотрено возможное решение для снижения количества ручных операций, выполняемых оператором системы за счет частичной автоматизации контроля вводимой информации. Решение основывается на использовании статистических критериев для обнаружения подозрительных значений (грубых выбросов). Рассмотренные критерии применимы к выборкам данных с нормальным законом распределения. В связи с этим предварительно выполнялась проверка гипотезы о соответствии распределения контролируемых величин этому закону.

Ключевые слова: мониторинг показателей, критерий Ирвина, критерий 3 сигма, критерий Шовене, критерий Граббса, критерий Гири, модифицированный критерий Колмогорова-Смирнова, модифицированный критерий хи-квадрат.

D.P. Khalitov¹, E.I. Molchanova¹

¹ *Irkutsk State Transport University, Irkutsk, Russian Federation*

ASSESSMENT OF THE POSSIBILITY TO REDUCE THE NUMBER OF MANUAL OPERATIONS TO PREPARE REPORTS AT IRKUTSK COMPUTING CENTER

Abstract. This article provides a review of the reporting generation system operated by the Irkutsk Computing Center, and also considers a possible solution for reducing the number of manual operations performed by the system operator through partial automation of control input information. The solution is based on the use of statistical criteria to detect suspicious values (gross outliers). The considered criteria are applicable to data samples with a normal distribution law. In this regard, a preliminary test of the hypothesis about the correspondence of the distribution of controlled quantities to this law was carried out.

Keywords: indicator monitoring, Irvin's criterion, 3 sigma test, Chauvenet's criterion, Grubbs' test, Geary's criterion, modified Kolmogorov-Smirnov criterion, modified Chi-square test.

Введение. Сотрудники Иркутского ИВЦ ежедневно в целях обеспечения специалистов и руководителей дороги справочным материалом формируют порядка 30 документов, каждый из которых содержит от 1 до 96 листов. Данные документы должны быть выданы точно к установленному времени, т.к. в дальнейшем используются на оперативных совещаниях. Кроме того, на основе показателей в этих документах принимаются управленческие решения, поэтому недопустимо предоставление недостоверной информации. В связи с этим перед публикацией документы должны быть тщательно проверены, очевидные замечания устранены, а по сомнительным значениям должны быть готовы пояснения причин.

Учитывая важность формируемой отчетности, технологи после формирования документов осуществляют визуальный контроль значений показателей. Такую работу сотрудник может проводить предварительно накопив необходимый опыт. Причины недостоверных показателей могут быть вызваны сбоями в работе программно-технических комплексов или человеческим фактором.

Целью данной статьи является обзор эксплуатируемой Иркутским ИВЦ информационной системы формирования отчетности и рассмотрение возможности снижения количества ручных операций, выполняемых оператором. Для достижения поставленной цели потребовалось выполнить обзор функций системы и ее компонентов.

Информационная система «Оперативная отчетность» и ее компоненты. Система «Оперативная отчетность» предназначена для формирования и предоставления пользователям Восточно-Сибирской железной дороги отчетных документов, отражающих специфику

территориального управления в разрезах выполнения основных эксплуатационных показателей работы железнодорожного транспорта, на основе данных из автоматизированных систем ОАО «РЖД».

Основные функции системы включают в себя ведение и хранение показателей работы железной дороги, выдача необходимых аналитических документов пользователям дороги, а также реализация удобного интерфейса для создания шаблонов выходных документов. В состав системы входят компоненты, важнейшими из которых являются: приложение ввода информации (scan_box.exe), веб-портал, приложение для формирования отчетных документов (index_auto.xlsm), приложение для ручного ввода данных (АРМ Макет), микросервисы «Load_maket_to_db.php» и «Proverka_polnoty.php».

Перед тем как приступить к дальнейшему описанию системы представим общую схему информационных потоков в нотации DFD (рис. 1).

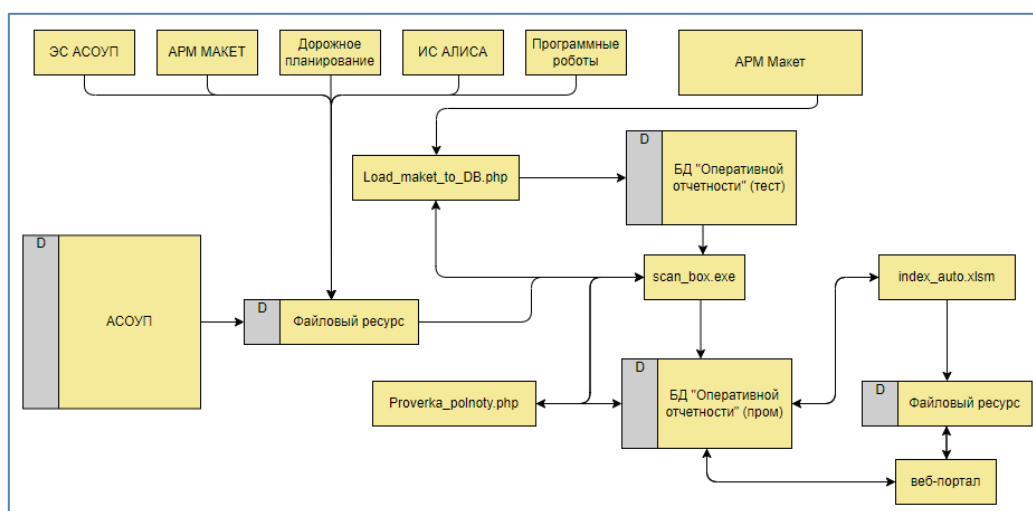


Рис. 1. Схема информационных потоков ИС «Оперативная отчетность»

Как видно на схеме, хранение данных организовано в двух базах данных - тестовой (БД «Оперативной отчетности» (тест)) и основной (БД «Оперативной отчетности» (пром)). Это связано с планами дальнейшего развития информационной системы, а также соображениями информационной безопасности. Далее рассмотрим каждую из подсистем подробнее.

Приложение для ввода информации в базу данных (scan_box.exe) представляет собой десктопное приложение, реализованное с использованием языка C#. Основной источник информации для «Оперативной отчетности» – это **сообщения** из автоматизированной системы оперативного управления перевозками (АСОУП), также называемые **макетами**. Они представляют собой структурированные сообщения – текстовые конструкции, создаваемые по определенным правилам. Сообщения имеют признаки начала «(:» и конца «:»)», отсутствие любого из них расценивается как ошибка. Структура макета также определяется такими понятиями как фразы и поля, которые условно можно назвать строками и столбцами.

Фраза – группа символьных значений, которая начинается с двоеточия. Первая фраза сообщения, содержащая данные идентифицирующие макет, называется служебной, в ней содержится следующая информация: код макета, код пункта передачи (КПП), дата, уровень раздельного учета (СНГ, собственные, ОАО РЖД, аренда и т.д.). В зависимости от своей структуры макеты делятся на позиционные и координатные. В позиционном макете каждая фраза имеет постоянное количество полей, отсутствующее значение воспринимается как ошибка. В координатном макете фраза имеет вид: [Координата1] <пробел> [Значение1] <пробел> [Координата2] <пробел> [Значение2] <пробел> ... <пробел> [КоординатаN] <пробел> [ЗначениеN].

Всего в системе зарегистрировано 168 макетов. Для каждого макета в БД отведена отдельная таблица, в связи с этим схему базы данных (БД) здесь не приводим.

Приложение для ввода информации в базу данных (scan_box.exe) позволяет выполнять следующие функции: вводить в БД «Оперативной отчетности» информацию из других систем, осуществлять форматный контроль вводимой информации, контролировать полноту поступления макетов для каждого документа, сравнивать поступивший макет с ранее введенным для выявления различий, просматривать уже введенные в БД макеты, удалять введенные в БД макеты.

При запуске программа ввода информации начинает сканировать специальный каталог, куда поступают макеты из АСОУП. Обнаружив файл, программа передает его содержимое микросервису «Load_maket_to_db», который осуществляет проверку по следующим параметрам: корректность структуры файла (количество полей, фраз и т.д.), прописан ли в БД номер сообщения/КПП/уровень учета, соответствует ли указанная в сообщении дата отчетным суткам, являются ли указанные в сообщении значения корректными числовыми величинами. При положительном результате проверки сервис «Load_maket_to_db» загружает сообщение в таблицу тестовой базы данных. В случае выявления ошибок в файле – в приложение scan_box.exe возвращается код ошибки, на основе которого программа информирует оператора о выявленных замечаниях, а сам файл перемещается в специальный каталог для проведения дальнейшего анализа. После загрузки сообщений из текстовых файлов, программа проверяет наличие новых макетов в тестовой базе данных и при обнаружении производит их перекачку в основную базу данных.

При проверке полноты информации программа scan_box.exe обращается к микросервису «Proverka_polnoty», который составляет список поступивших макетов и сравнивает его с перечнем макетов, необходимых для формирования документов. В результате оператор в оперативном режиме контролирует готовность базы данных к формированию каждого документа.

Веб-приложение для просмотра отчетов служит для организации доступа пользователей к сформированным отчетам, учитывая отраслевые требования информационной безопасности. Оно реализовано с использованием языков PHP, JavaScript и работает под управлением веб сервера Microsoft IIS. Установка на клиентский компьютер не требуется. После включения пользователя в соответствующую группу Active Directory, на рабочем столе его компьютера появляется ярлык, с помощью которого можно открыть начальную страницу и работать с документами.

Кроме организации просмотра отчетных документов, веб-приложение позволяет настроить персональный дашборд – табло, визуализирующее динамику изменений выбранных пользователем показателей работы дороги (рис. 2).

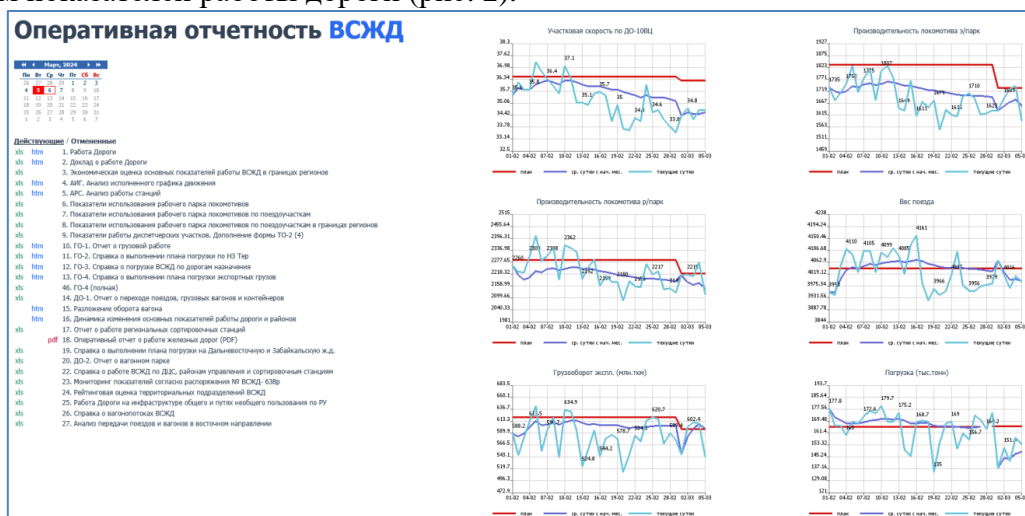


Рис. 2. Главная страница веб приложения «Оперативная отчетность».

Программа формирования отчетных документов (index_auto.xlsm) реализована с использованием табличного процессора MS Excel и является XLSM файлом, в котором содержатся макросы, реализованные на языке Visual Basic for Automation. Эта программа позволяет

создавать документы в форматах XLS и PDF, автоматически копировать их в соответствующие каталоги на веб-сервере и заполнять таблицы в базе данных, которые используются для отображения пользователям веб-версий документов.

Основные функции программы формирования документов: формирование отчетных документов, информирование пользователя о ходе процесса формирования документов, формирование данных для смежных систем (данные для АРМ технического нормирования и др.). Формирование выходных отчетов в «Оперативной отчетности» происходит с использованием заранее подготовленных технологами шаблонов документов. Данные шаблоны представляют собой файлы формата XLS, размеченные по определенным правилам. Ключевым понятием в разметке документов является служебная фраза, которая вносится в ячейки MS Excel в следующем виде: ="@|xxx1|xxx2|xxx3|xxx4|xxx5|xxx6|xxx7", где:

xxx1 – информация о дате. Некоторые допустимые значения: «пл» - первое числа текущего месяца, «тс» - отчетные сутки текущего месяца, «нсм» - накопление с начала месяца, «мсс» - средние сутки с начала месяца, «нстг» - накопление с начала года и т.п.

xxx2 – номер сообщения (макета)

xxx3 – код пункта передачи (КПП). Допускается перечислять КПП через запятую. В этом случае значения из сообщений по перечисленным пунктам передач суммируются.

xxx4 – номер фразы в сообщении (макете). Допускается перечислять фразы через запятую. В этом случае значение из сообщения по перечисленным фразам суммируются.

xxx5 – номер поля в сообщении (макете)

xxx6 – уровень учета. Допускается перечислять уровни учета через запятую. В этом случае значения из сообщений по перечисленным уровням учета суммируются.

xxx7 – тип макета. Допустимые значения: 0 – координатный, 1 – позиционный.

Пример служебной фразы: ="@|пл|6271|1,2,3,4|20,40,60|1_1|20,200|0"

С помощью макросов в программе формирования документов осуществляется «проход» по определенным листам и диапазонам ячеек шаблона документа. Встречая служебную фразу, макрос: производит ее разбор, и, в зависимости от содержания, формирует SQL запрос; отправляет SQL запрос к базе данных; подменяет служебную фразу полученным из базы данных значением. Важно отметить, что запуск формирования документов автоматизирован и происходит с учетом определенных временных рамок и при условии достижения полноты необходимой информации.

Технологическая схема формирования документов в ИС «Оперативная отчетность» представлена на рис. 3.

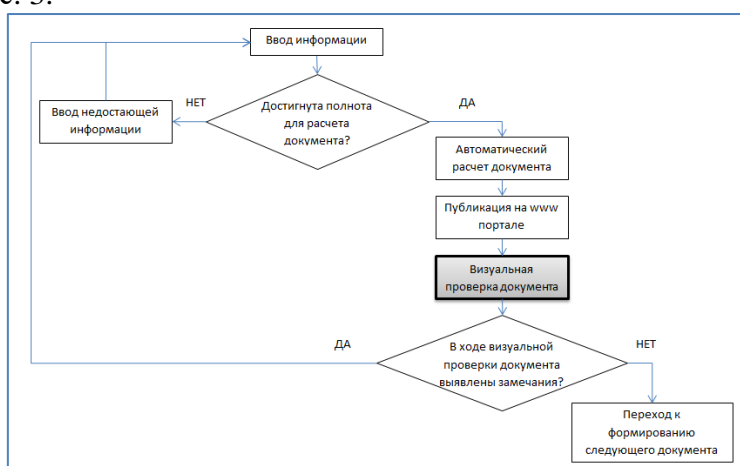


Рис. 3. Технологическая схема формирования документов в ИС «Оперативная отчетность»

Приложение для ручного ввода данных. АРМ Макет – это веб приложение, предоставляющее пользователю интерфейс для ввода данных. Оно реализовано с использованием языков PHP, JavaScript и работает под управлением веб сервера Microsoft IIS. Важной особенностью приложения «АРМ Макет» является то, что с помощью специального редактора технолог может создать в нем любое сообщение, описать и настроить его структуру, выделить права

пользователям для доступа к данному сообщению. Все эти операции осуществляются без привлечения программистов.

Для формализации ввода макетов пользователю на экран выводится таблица (или таблицы). Информация, введенная пользователем в таблицу (таблицы), не идентична по структуре сообщению, которое передается в базу данных «Оперативной отчетности». Поэтому АРМ Макет позволяет также создавать, редактировать, удалять координатные группы – набор фраз и полей, которые отправляются в систему-приемник.

Т.е. таблицы предназначены для ввода сообщений в клиентском приложении. Их количество и состав могут не совпадать с количеством и составом координатных групп. Например, координатные группы с большим количеством полей для удобства ввода можно разбить на две таблицы, или наоборот – две координатные группы объединить в одну таблицу.

Для гибкой настройки проверки вводимой информации АРМ Макет позволяет указывать маску фразы – специальная строка, определяющая формат каждой ячейки в таблице. Маска может содержать символы:

* - данная ячейка блокируется на запись для всех пользователей; % - разделитель дробной части не передается, к введенному числу применяется маска, а затем разделитель удаляется; + - число должно быть строго положительным.

Вышеперечисленные символы должны стоять в самом начале маски в той последовательности, в которой они перечислены.

- необязательная цифра (если не введена, то не передается); 0 - обязательная цифра (если не введена, то передается 0); «.», «,» - точка или запятая - разделитель дробной части.

Если введенное значение после запятой содержит больше символов, чем количество знаков после запятой в маске, то число округляется. Если же переполнение происходит в целой части, то выдается сообщение об ошибке. Маска может так же содержать любые другие символы (кроме перечисленных выше и «;»), которые передаются в текст сообщения без изменения. Если маска для какой-нибудь ячейки не найдена или признана синтаксически ошибочной, то она принимается равной #####0.####.

Формат строки с масками ячеек: «код_поля1=маска1; код_поля2=маска2; ...».

Возможное решение по оптимизации количества ручных операций, выполняемых оператором ИС «Оперативная отчетность». Изучив функции рассматриваемой информационной системы, можно отметить, что данная система имеет высокий уровень автоматизации работы оператора. Ввод данных, запуск формирования документов и сам процесс формирования документов не требует вмешательства технолога. За человеком остается только контроль работы системы и устранение аварийных ситуаций в случае сбоев. Но все-таки не все этапы подготовки отчетности автоматизированы. Как уже было сказано выше, проверка поступающей информации производится технологом (серый блок на рис. 3). Причем эта работа осуществляется уже после формирования документов, т.к. проверять поступающую информацию в виде сообщений весьма трудоемко, учитывая объемы информации в сообщениях и количество самих сообщений.

В качестве возможного решения автоматизации контроля вводимой информации мы предлагаем применить статистические критерии для обнаружения грубых выбросов. В математической статистике известно достаточно много методов, позволяющих выявлять аномальные значения, не согласующиеся с остальными элементами выборки. Например: критерий Граббса [1-2], критерий Ирвина [3], критерий 3-сигма [3,4], критерий Романовского [3,5], критерий Шовене [6-8], критерий Пирсона [9]. Использование этих критериев предполагает принадлежность наблюдаемых случайных величин закону нормального распределения. При этом вопрос выбора критериев, позволяющих проверить гипотезу о принадлежности наборов данных нормальному закону, не является простым. В [10] отмечено, что «не смотря на множество публикаций, не хватает объективной информации о действительных свойствах критериев, их достоинствах и недостатках ... какие критерии предпочтительнее использовать и почему».

В работе [11] нами была приведена оценка возможности обнаружения подозрительных значений в суточных показателях работы с помощью критериев Ирвина, 3-сигма, Шовене,

Грabbca. Проверить любой массив одномерных данных с помощью этих критериев можно используя специальное программное обеспечение, реализованное нами с использованием языка Python.

База данных ИС «Оперативная отчетность» хранит в себе множество различных показателей детализируемых по разнообразным параметрам (дата, территориальный объект дороги, род подвижного состава, уровень учета, вид тяги и т.д.). Для практической апробации предложенного решения было взято 10 показателей работы дороги из базы данных «Оперативной отчетности», представленных в таблице 2, за период с 01.01.2018 г по 28.03.2024 г.

Таблица 2.

Перечень исследуемых показателей

№ показателя	Наименование показателя	Единицы измерения	Диапазон
1.	Прием вагонов	Вагоны	4211 - 14958
2.	Сдача вагонов	Вагоны	6032 - 13782
3.	Маршрутная скорость грузового поезда	км/сут	357,65 - 850,16
4.	Оборот грузового вагона	Сутки	5,13 - 11,74
5.	Процент порожнего пробега	%	23,36 - 46,42
6.	Производительность вагона	ткм/ваг	5360,16 - 12893,95
7.	Парк порожних вагонов	Вагоны	28454 - 43511
8.	Парк Местных вагонов	Вагоны	3237 – 10860
9.	Парк Транзитных вагонов	Вагоны	13943 – 27653
10.	Парк Неисправных вагонов	Вагоны	2058 – 4300

Результаты работы программы представлены ниже в таблицах 3 - 5.

Таблица 3.

Результаты тестов на соответствие нормальному распределению до исключения грубых выбросов

Проверка гипотезы о принадлежности нормальному закону критерием Гир										
	1	2	3	4	5	6	7	8	9	10
1 мес	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	-	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	-	+	+	+	+	-	-	+
12 мес	-	+	-	+	+	-	-	-	+	-
Весь период	+	+	+	-	-	+	-	-	-	+
Проверка гипотезы о принадлежности нормальному закону мод. кр. Колмогорова-Смирнова										
	1	2	3	4	5	6	7	8	9	10
1 мес	+	+	+	+	+	+	+	+	-	+
2 мес	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	+	-	+	+	+	-	-	+
12 мес	-	-	-	+	+	+	-	-	+	+
Весь период	-	-	-	-	-	-	-	-	-	-
Проверка гипотезы о принадлежности нормальному закону критерием χ^2										
	1	2	3	4	5	6	7	8	9	10
1 мес	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-
6 мес	+	+	+	+	+	+	+	-	-	+
12 мес	+	+	+	+	+	+	+	-	+	-
Весь период	-	-	-	-	-	-	-	-	-	-

Таблица 4.

Результаты работы программы по выявлению грубых выбросов

Тест Ирвина

	1 мес	2 мес	3 мес	6 мес	12 мес	Весь период
Показатель № 1:						4211, 4250
Показатель № 2:						
Показатель № 3:						
Показатель № 4:						11.74
Показатель № 5:						23.36
Показатель № 6:		5451.35	5451.35	5451.35	5451.35	
Показатель № 7:						
Показатель № 8:						
Показатель № 9:						
Показатель № 10:						
Тест 3 сигма						
	1 мес	2 мес	3 мес	6 мес	12 мес	Весь период
Показатель № 1:						4211, 4250, 5509, 5557, 6298, 6504, 7061, 7953, 8031, 14229, 14264, 14958
Показатель № 2:				8381.0	7495.0	6032, 6043, 6612, 7049, 7495, 7776, 7788, 7916, 7921, 7948
Показатель № 3:						357.65, 839.71, 840.13, 844.94, 849.71, 850.16
Показатель № 4:		7.57	7.57	7.99, 8.19	8.44	8.46, 8.48, 8.48, 8.57, 8.62, 8.63, 8.84, 8.86, 9.01, 9.38, 9.4, 9.45, 9.56, 9.86, 10.48, 11.74
Показатель № 5:					28.47	23.36, 27.06, 27.29, 28.1, 46.42
Показатель № 6:		5451.35	5451.35	5451.35	5451.35	5360.16, 5415.69, 5451.35, 5491.2, 12412.88, 12563.15, 12597.63, 12742.85, 12855.02, 12885.95, 12893.95
Показатель № 7:				37764.0, 38220.0	38220.0	42168, 42171, 42240, 42254, 42332, 42597, 42914, 43303, 43511
Показатель № 8:						10545, 10558, 10601, 10760, 10814, 10860
Показатель № 9:						26743, 26747, 26883, 26935, 27572, 27653
Показатель № 10:						4279, 4285, 4288, 4300, 4300
Тест Шовене						
	1 мес	2 мес	3 мес	6 мес	12 мес	Весь период
Показатель № 1:	9411.0					4211.0, 4250.0, 5509.0, 5557.0, 6298.0, 6504.0, 7061.0, 14958.0
Показатель № 2:				8381.0	7495.0	6032.0, 6043.0, 6612.0, 7049.0
Показатель № 3:	705.59	705.59				
Показатель № 4:	7.57	7.57	7.57	7.99, 8.19	8.44	9.01, 9.38, 9.4, 9.45, 9.56, 9.86, 10.48, 11.74
Показатель № 5:	31.97	31.97, 40.77	41.05, 41.22		28.47	23.36
Показатель № 6:		5451.35	5451.35	5451.35	5451.35	5360.16
Показатель № 7:				37764.0, 38220.0		
Показатель № 8:	5702.0					
Показатель № 9:						
Показатель № 10:		3488.0				
Тест Граббса						
	1 мес	2 мес	3 мес	6 мес	12 мес	Весь период
Показатель № 1:						4211, 4250, 5509, 5557, 6298, 6504
Показатель № 2:						6032.0, 6043.0
Показатель № 3:						

Показатель № 4:											9.38, 9.4, 9.45, 9.56, 9.86, 10.48, 11.74
Показатель № 5:											23.36
Показатель № 6:			5451.35	5451.35	5451.35						
Показатель № 7:											
Показатель № 8:											
Показатель № 9:											
Показатель № 10:											

Таблица 5.

Результаты тестов на соответствие нормальному распределению после исключения грубых выбросов

NormalityTest_Geary(искл_кр_Ирвина):											NormalityTest_K-Smirnov(искл_кр_Ирвина):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	+	+	1 мес	+	+	+	+	+	+	+	+	+	-	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	-	+	+	-	+	-	-	+	6 мес	-	+	+	-	+	+	+	-	-	+	+
12 мес	-	+	-	+	+	-	-	-	+	-	12 мес	-	-	-	+	+	+	-	-	+	+	+
Все время	+	+	+	-	-	+	-	-	-	+	Все время	-	-	-	-	-	-	-	-	-	-	-
NormalityTest_chi square(искл_кр_Ирвина):											NormalityTest_Geary(искл_кр_3 сигма):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	+	+	1 мес	+	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	+	+	+	+	+	-	+	-	-	+	6 мес	-	+	-	+	+	-	+	-	-	+	+
12 мес	+	+	+	+	+	+	+	-	+	-	12 мес	-	+	-	+	+	-	-	-	+	+	-
Все время	-	-	-	-	-	-	-	-	-	-	Все время	-	+	+	+	-	-	-	-	-	+	+
NormalityTest_K-Smirnov(искл_кр_3 сигма):											NormalityTest_chi square(искл_кр_3 сигма):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	-	+	1 мес	+	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	+	-	+	+	+	-	-	+	6 мес	+	+	+	-	+	-	+	-	-	+	+
12 мес	-	-	-	+	+	+	-	-	+	+	12 мес	+	+	+	+	+	+	+	-	+	+	-
Все время	-	-	-	-	-	-	-	-	-	-	Все время	-	-	-	-	-	+	-	-	-	-	-
NormalityTest_Geary(искл_кр_Шовене):											NormalityTest_K-Smirnov(искл_кр_Шовене):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	+	+	1 мес	+	+	+	+	+	+	+	+	+	-	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	-	+	+	-	+	-	-	+	6 мес	-	+	+	-	+	+	+	-	-	+	+
12 мес	-	+	-	+	+	-	-	-	+	-	12 мес	-	-	-	+	+	+	-	-	+	+	+
Все время	-	+	+	+	-	+	-	-	-	+	Все время	-	-	-	-	-	-	-	-	-	-	-
NormalityTest_chi square(искл_кр_Шовене):											NormalityTest_Geary(искл_кр_Граббса):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	+	+	1 мес	+	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	+	+	+	-	+	-	+	-	-	+	6 мес	-	+	-	+	+	-	+	-	-	+	+
12 мес	+	+	+	+	+	+	+	-	+	-	12 мес	-	+	-	+	+	-	-	-	+	+	-
Все время	-	-	-	-	-	-	-	-	-	-	Все время	-	+	+	-	-	+	-	-	-	-	+
NormalityTest_K-Smirnov(искл_кр_Граббса):											NormalityTest_chi square(искл_кр_Граббса):											
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10	
1 мес	+	+	+	+	+	+	+	+	-	+	1 мес	+	+	+	+	+	+	+	+	+	+	+
2 мес	+	+	+	+	+	+	+	+	+	+	2 мес	+	+	+	+	+	+	+	+	+	+	+
3 мес	+	+	+	+	+	+	+	+	+	-	3 мес	+	+	+	+	+	+	+	+	+	+	-
6 мес	-	+	+	-	+	+	+	-	-	+	6 мес	+	+	+	+	+	-	+	-	-	+	+

12 мес	-	-	-	+	+	+	-	-	+	+	12 мес	+	+	+	+	+	+	-	+	-
Все время	-	-	-	-	-	-	-	-	-	-	Все время	-	-	-	-	-	-	-	-	-

Как мы видим, исключение грубых выбросов из наборов данных слабо влияет на результаты проверки гипотезы о нормальности распределения. Это может говорить о том, что модифицированный критерий Колмогорова-Смирнова, χ^2 и критерий Гири «не реагируют» на исключение выбросов, вероятно, в связи с большим объемом исследуемых выборок, в которых в разные периоды времени у показателей может меняться характер поведения - проявляться сезонность, цикличность или тренд. Т.е. вопрос выбора периода для формирования исследуемых наборов данных требует дальнейшей проработки.

Тем не менее, исходя из результатов, представленных в Таблице 4, можно утверждать, что автоматизация проверки сформированных документов на предмет обнаружения подозрительных значений в суточных показателях возможна.

Критерии обладают разной «чувствительностью» к грубым выбросам, поэтому решение об отнесении значения к выбросам необходимо применять после применения нескольких критериев выявления грубых выбросов, что подтверждает вывод сделанный Орловым А.И. на основе теории устойчивости [12].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Grubbs F.E. Sample Criteria for Testing Outlying observations // Annals of Mathematical Statistics. 1950. V. 63, № 1. P. 27-58.
2. Grubbs F.E. Procedures for Detecting Outlying Observations in Samples // Technometrics. 1969. V. 11, №. 1. P. 1-21.
3. Третьяк Л. Н. Обработка результатов наблюдений: Учебное пособие. Оренбург: ГОУ ОГУ, 2004. 171 с.
4. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. Л.: Энергоатомиздат, 1991. 304 с.
5. Марчук В.И., Токарева С.В. Способы обнаружения аномальных значений при анализе нестационарных случайных процессов. Шахты.: ЮРГУС, 2009. 209 с.
6. Тейлор Дж. Введение в теорию ошибок. Пер. с англ.— М.: Мир, 1985. —272 с.
7. Большев Л.Н., Убайдуллаева М. Критерий Шовенэ в классической теории ошибок // Теория вероятностей и ее применение. 1974. Т. XIX, вып. 4. С. 714-723.
8. Пагурова В.И. О критерии Шовэнэ для обнаружения нескольких выбросов // Теория вероятностей и ее применение. 1985. Т. 30, № 3. С. 558-561.
9. Pearson E.S., Chandra S.C. The efficiency of statistical tools and a criterion for rejection of outlying observations // Biometrika. 1936. V. 28, № 3/4. P. 308-320.
10. Б.Ю. Лемешко Критерии проверки отклонения распределения от нормального закона. Руководство по применению. // Новосибирский государственный технический университет (2014)
11. Khalitov, Dmitry & Molchanova, Elena. (2024). An approach to automating the daily performance indicators monitoring of a transport company. E3S Web of Conferences. 471. 10.1051/e3sconf/202447104024.
12. Орлов А.И. Ошибки при использовании коэффициентов корреляции и детерминации. Заводская лаборатория. Диагностика материалов. 2018;84(3):68-72.
13. А.И. Кобзарь, Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
14. Лемешко Б.Ю. О проблемах и ошибках применения критериев согласия // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 74–90.
15. Я.М. Радкевич, Метрология, стандартизация и сертификация в 3 ч. Часть 1. Метрология: учебник для среднего профессионального образования / Я. М. Радкевич, А. Г. Схиртладзе. — 5-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 235 с.

16. Keya Rani Das, A. H. M. Rahmatullah Imon. A Brief Review of Tests for Normality. American Journal of Theoretical and Applied Statistics. Vol. 5, No. 1, 2016, pp. 5-12.

REFERENCES

1. Grubbs F.E. Sample Criteria for Testing Outlying observations // Annals of Mathematical Statistics. 1950. V. 63, № 1. P. 27-58.
2. Grubbs F.E. Procedures for Detecting Outlying Observations in Samples // Technometrics. 1969. V. 11, №. 1. P. 1-21.
3. Tret'yak L. N. Obrabotka rezul'tatov nablyudenij: Uchebnoe posobie. Orenburg: GOU OGU, 2004. 171 s.
4. Novickij P.V., Zograf I.A. Ocenka pogreshnostej rezul'tatov izmerenij. L.: Energoatomizdat, 1991. 304 s.
5. Marchuk V.I., Tokareva S.V. Sposoby obnaruzheniya anomal'nyh znachenij pri anali-ze nestacionarnyh sluchajnyh processov. SHahty.: YURGUS, 2009. 209 s.
6. Tejlor Dzh. Vvedenie v teoriyu oshibok. Per. s angl.— M.: Mir, 1985. —272 s.
7. Bol'shev L.H., Ubajdullaeva M. Kriterij SHovene v klassicheskoj teorii oshibok // Teoriya veroyatnostej i ee primenenie. 1974. T. XIX, vyp. 4. S. 714-723.
8. Pagurova V.I. O kriterii SHovene dlya obnaruzheniya neskol'kih vybrosov // Teoriya veroyatnostej i ee primenenie. 1985. T. 30, № 3. S. 558-561.
9. Pearson E.S., Chandra S.C. The efficiency of statistical tools and a criterion for rejection of outlying observations // Biometrika. 1936. V. 28, № 3/4. P. 308-320.
10. B.YU. Lemeshko Kriterii proverki otkloneniya raspredeleniya ot normal'nogo zakona. Rukovodstvo po primeneniyu. // Novosibirskij gosudarstvennyj tekhnicheskij universitet (2014)
11. Khalitov, Dmitry & Molchanova, Elena. (2024). An approach to automating the daily performance indicators monitoring of a transport company. E3S Web of Conferences. 471. 10.1051/e3sconf/202447104024.
12. Orlov A.I. Oshibki pri ispol'zovanii koefficientov korrelyacii i determinacii. Zavodskaya laboratoriya. Diagnostika materialov. 2018; 84(3):68-72.
13. A.I. Kobzar', Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnyh rabotnikov. – M.: FIZMATLIT, 2006. – 816 s.
14. Lemeshko B.YU. O problemah i oshibkah primeneniya kriteriev soglasiya // Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika. 2023. № 64. S. 74–90.
15. YA.M. Radkevich, Metrologiya, standartizaciya i sertifikaciya v 3 ch. CHast' 1. Metrologiya: uchebnyk dlya srednego professional'nogo obrazovaniya / YA. M. Radkevich, A. G. Skhirtladze. — 5-e izd., pererab. i dop. — Moskva : Izdatel'stvo YUrajt, 2023. — 235 s.
16. Keya Rani Das, A. H. M. Rahmatullah Imon. A Brief Review of Tests for Normality. American Journal of Theoretical and Applied Statistics. Vol. 5, No. 1, 2016, pp. 5-12.

Информация об авторах

Дмитрий Павлович Халитов – магистрант, Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: usercom@inbox.ru

Елена Ивановна Молчанова – д. т. н., профессор, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: molchanova_ei@irgups.ru

Authors

Dmitry Pavlovich Khalitov – master's student, Irkutsk State Transport University, Irkutsk, e-mail: usercom@inbox.ru

Elena Ivanovna Molchanova, Doctor of Technical Science, Professor, the Subdepartment Information systems and information security, Irkutsk State Transport University, Irkutsk, e-mail: molchanova_ei@irgups.ru

Для цитирования

Халитов Д. П., Молчанова Е. И. Оценка возможности снижения количества ручных операций при эксплуатации системы формирования отчетности в Иркутском ИВЦ // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №1. – С.1-11 – Режим доступа <https://ismm.irkgups.ru/toma/121-2024>. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 17.04.2024)

For citations

Khalitov D.P., Molchanova E.I. Assessment of the possibility to reduce the number of manual operations to prepare reports at Irkutsk Computing Center // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: elektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal]. – 2024. – №1. – P.1-11 – Accessed: <https://ismm.irkgups.ru/toma/121-2024>.