

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ ЖЕЛЕЗНОДОРОЖНЫХ ПУТЕЙ

В. А. Голуб, Г. А. Поляков, А. А. Богданов, Е. В. Громоткова, О. А. Плишкина

Воронежский государственный университет

В статье предложены алгоритмы формирования репрезентативных выборок, необходимых для корректного статистического анализа состояния железнодорожных путей, а так же методика выявления зависимостей между дефектами железнодорожных путей и различными факторами, влияющими на появление и развитие дефектов.

Безопасность пассажиров и сохранность перевозимых на железнодорожном транспорте грузов определяется состоянием железнодорожных путей.

Оценка состояния железнодорожных путей осуществляется, прежде всего, на основании результатов измерений, регулярно проводящихся вагонами-путеизмерителями.

Однако до настоящего времени отсутствует корректный математический анализ этих данных и не выявлено какие именно факторы и насколько сильно влияют на возникновение и развитие дефектов железнодорожных путей.

В данной работе рассматривается алгоритм формирования репрезентативных выборок, необходимых для корректного статистического анализа состояния железнодорожных путей, и на основе применения аппарата математической статистики анализируются зависимости между дефектами железнодорожных путей и факторами, влияющими на эти дефекты, а также рассматривается программное обеспечение, разработанное для решения этих задач.

Сложность формирования выборок заключается в том, что различные факторы, влияющие на развитие дефектов, могут быть взаимно зависимы. Для того, чтобы корректно исследовать влияние каждого из факторов на появление того или иного дефекта пути, необходимо зафиксировать все остальные факторы и рассмотреть различные комбинации анализируемого фактора и дефектов путей, т. е. пары (D_i, F_j) , $i = 1..6$.

Основными дефектами железнодорожных путей являются: сужение (D_1), уширение (D_2), уровень (D_3), перекос (D_4), просадка (D_5), рихтовка (D_6). Измерение этих дефектов осуществляется вагоном-путеизмерителем [4].

© Голуб В. А., Поляков Г. А., Богданов А. А., Громоткова Е. В., Плишкина О. А., 2006

Основными факторами, влияющими на появление дефектов железнодорожных путей, являются: тип скрепления (F_1) — деревянные, железобетонные или анкерные шпалы, грузонапряженность (F_2), план линии (F_3), продольный профиль пути (F_4), пропущенный тоннаж (F_5), установленная скорость движения (F_6) [4].

Для исследования зависимости той или иной неисправности пути от некоторого фактора, например, плана пути, необходимо формировать выборку таким образом, чтобы все остальные факторы, влияющие на неисправность (продольный профиль, установленная скорость, пропущенный тоннаж, тип скрепления, вес поезда, количество поездов в сутки) были по возможности одинаковыми. При этом формировать выборку следует так, чтобы ее объем был максимально возможным. Несмотря на значительное число измерений, выполненных на разных участках дороги вагоном-путеизмерителем, удовлетворить одновременно обоим требованиям оказывается крайне не просто.

Для решения задачи формирования репрезентативных выборок можно предложить следующие два подхода: через нахождение δ -окрестности среднего значения факторов на участке пути или через построение гистограммы частот выборочных данных. Рассмотрим эти методы.

МЕТОД ФОРМИРОВАНИЯ ВЫБОРКИ НА ОСНОВЕ ОПРЕДЕЛЕНИЯ δ -ОКРЕСТНОСТИ СРЕДНЕГО КОЛИЧЕСТВА ЗНАЧЕНИЙ ФАКТОРОВ НА УЧАСТКЕ ПУТИ

Предлагаемый алгоритм формирования выборки выполняется в несколько этапов.

1. Выбирается N участков пути. Каждый из участков содержит n километров, для которых

выполнены измерения параметров пути вагоном-путеизмерителем и имеются данные о факторах, влияющих на появление и развитие дефектов пути.

2. Для k -ого участка пути ($k = 1, \dots, N$) определяется среднее значение анализируемого фактора \bar{x}_j по формуле

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad (1)$$

где $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ — выборочные значения случайной величины ξ_j , т.е. x_{ij} — общее число значений j -ого фактора на i -м километре выбранного участка пути, $i = \overline{1, \dots, n}$; n — объем выборки, равный числу километров на рассматриваемом участке.

3. Строится δ -окрестность полученного значения. При условии, что величина d_k исследуемого дефекта известна, для k -ого участка имеем пару выборочных значений (x_{kj}, d_k) , где $x_{kj} \in (\bar{x}_{kj} - \xi, \bar{x}_{kj} + \xi)$, d_k — выборка соответствующего дефекта железнодорожного пути D_k . Поскольку имеется N участков, то получаем выборку объема N .

В качестве критерия допустимого отклонения δ от среднего значения выбирается значение 5—10 %, вполне удовлетворяющее допустимым для практики погрешностям.

МЕТОД ФОРМИРОВАНИЯ ВЫБОРКИ НА ОСНОВЕ ПОСТРОЕНИЯ ГИСТОГРАММЫ ЧАСТОТ ВЫБОРОЧНЫХ ДАННЫХ

Всю числовую ось разобьем на промежутки Δ_k , $k = \overline{1..N}$ так, что $R = \bigcup_k \Delta_k$ и $\Delta_k \cap \Delta_i = O$ для $k \neq i$. Рассмотрим частоту возникновения дефекта

$$v_k(\xi) = \sum_{i=1}^n I(x_{ij} \in \Delta_k) \quad (2)$$

как случайную величину, равную количеству элементов выборки, попавших в интервал Δ_k . Гистограмма описывается соотношением:

$$f_n^*(\xi, x_j) = \frac{1}{n} \sum_k \frac{v_k(\xi)}{|\Delta_k|} I_{\Delta_k}(x_j), \quad (3)$$

где $x_j \in R$ [1].

Данный алгоритм выполняется в шесть этапов:

1) определяется размах выборки ω , равный разности максимального и минимального значения исследуемого фактора;

2) находится длина интервала $\Delta_k = \frac{\omega}{N}$, где N — число интервалов разбиения;

3) по формуле (2) подсчитываются частоты v_k — число элементов наблюдения, попавших в интервал Δ_k , причем элемент, совпадающий с верхней границей интервала, относится к последующему интервалу;

4) определяется максимальная частота $v = \max(v_k)$;

5) выборка формируется из элементов исследуемого фактора, попадающих в интервал Δ , соответствующий максимальной частоте v ;

6) при условии, что величина исследуемого дефекта известна (d_i), имеем пару выборочных значений (x_{ij}, d_i) для i -ого участка, где $x_{ij} \in \Delta$.

Данный подход к формированию выборок позволяет по максимуму гистограммы определить наиболее часто встречающиеся значения фактора F_j , а значит и обеспечить наибольший объем выборки в случае, когда фиксированные значения близки.

Для получения репрезентативных выборок в соответствии с предложенными выше алгоритмами разработана программа формирования статистических данных о состоянии железнодорожных путей. Программа обеспечивает обработку информации о параметрах железнодорожных путей, получаемой вагоном-путеизмерителем и ее конвертацию в базу данных, входящую в состав программы. Указанная база данных содержит результаты измерений, характеризующие состояние железнодорожных путей, а также информацию о факторах, которые влияют на появление и развитие дефектов железнодорожных путей. Программа позволяет формировать выборки для каждого из дефектов (сужение, уширение, уровень, перекося, просадка, рихтовка) в зависимости от каждого из факторов (план пути, продольный профиль пути, установленная скорость, пропущенный тоннаж, тип скрепления) за определенный период времени, при этом фиксируется заданное пользователем направление и номер пути.

В соответствии с алгоритмом решения задачи схему обработки данных можно представить следующим образом (рис.):

Исследование зависимости определенного дефекта от заданного фактора осуществляется путем нахождения коэффициента корреляции для пары выборок, соответствующей рассматриваемым дефекту и фактору соответственно.



Рис. Схема обработки данных в программе формирования выборок и анализа состояния железнодорожных путей

Коэффициент парной корреляции находится по формуле [3]:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}. \quad (12)$$

Достоверность коэффициента корреляции оценивалась по критерию t -Стьюдента. Для этого необходимо рассчитать коэффициент t :

$$t = \sqrt{\frac{r^2 \cdot (n - 2)}{1 - r^2}}, \quad (13)$$

где $m_r = \sqrt{\frac{1 - r^2}{n - 2}}$ есть среднеквадратическая ошибка коэффициента корреляции [4].

Значение коэффициента t при, например, 5 % уровне значимости определяется по распре-

делению t -Стьюдента при $\alpha = 0.05$ и числе степеней свободы $k = n - 2$. Достоверность коэффициента корреляции подтверждается, когда $t_{кр} = t(k, \alpha) < t$ [4].

Разработанная программа исследования дефектов железнодорожных путей дает возможность исследовать зависимость каждого из дефектов (сужение, уширение, уровень, перекос, просадка, рихтовка) от каждого из факторов (план пути, продольный профиль пути, установленная скорость, пропущенный тоннаж). Это позволяет выяснить, какие именно неблагоприятные факторы наиболее существенно влияют на появление и развитие тех или иных дефектов железнодорожных путей.

Входными данными программы являются данные о количестве и величине дефектов (определенных по результатам измерений, выполненных вагоном-путеизмерителем) железнодорожных путей и факторах, которые могут влиять на состояние железнодорожной линии. На выходе программа выдает информацию о степени зависимости того или иного дефекта железнодорожного пути от данного фактора по величине коэффициента корреляции.

Как показывают расчеты, при анализе каждого из факторов, влияющих на появление дефектов железнодорожных путей, взятого отдельно от других факторов (даже при формировании выборки в соответствии с указанным выше алгоритмом, обеспечивающим малые изменения значений всех факторов, кроме анализируемого), ни один из рассматриваемых факторов не является определяющим для порождения того или иного дефекта. Например, величина коэффициента корреляции фактора «план пути» и дефекта «уровень» 3-й степени составляет 0,104, что указывает на слабую статистическую зависимость этих характеристик. Зависимость дефекта «посадка» 3-й степени от фактора «продольный профиль пути» является еще более слабой (коэффициент корреляции равен 0,047).

Предложенная в работе методика формирования выборок позволяет анализировать влияющие на появление дефектов пути факторы по отдельности. Полученные результаты позволяют сделать вывод об отсутствии среди проанализированных факторов доминирующих, т. е. таких, влияние которых на появление дефектов было бы определяющим, а значит, выделить

среди рассматриваемых факторов наиболее значимый по имеющимся данным не представляется возможным.

СПИСОК ЛИТЕРАТУРЫ

1. *Радченко Т.А.* Теория вероятностей и математическая статистика. / Т. А. Радченко, Ю. С. Радченко. — Воронеж: Изд-во ВГУ, 1997. — 240 с.

2. *Гмурман В.Е.* Теория вероятностей и математическая статистика. / В. Е. Гмурман. — М.: Высшая школа, 1997. — 479 с.

3. *Крамер Г.* Математические методы статистики / Г. Крамер; Пер. с англ. А. С. Мониной, А. А. Петрова; Под ред. акад. А. Н. Колмогорова. — М. : Мир, 1975. — 648 с.

4. Железные дороги: Общий курс: учебник для вузов ж.-д. трансп. / М. М. Филиппов, М. М. Уздин, Ю.И. Ефименко [и др.]; под ред. М. М. Филиппова. — М.: Транспорт, 1981. — 343 с.