

**ТЕХНОЛОГІЇ ТА
ОБЛАДНАННЯ ВИРОБНИЦТВА**

УДК 621.3:622:519.24

Б. Б. КОБЫЛЯНСКИЙ**ОЦЕНКА НЕКОТОРЫХ ОСОБЕННОСТЕЙ ГОРНОГО ПРОИЗВОДСТВА**

На современном этапе технического развития все комплексные системы промышленного производства и в частности угольной промышленности, требуют автоматизации всего процесса включающего работу добычного и транспортного оборудования, что нельзя выполнить без детального математического описания всех специфических явлений характеризующих конкретное производство. Поэтому, для решения задачи управления сложными технологическими, многофакторными системами, предлагается использование непараметрической статистики, которые в отечественной специальной литературе не нашли должного отражения.

Ключевые слова: угольная промышленность, оценка, непараметрическая статистика.

Введение. На современном этапе предприятия угольной промышленности оснащены автоматизированными линиями и комплексами, внедряются современные технологические комплексы и участки, гибкие производственные системы. Все это способствует решению двух взаимосвязанных задач: выпуск более качественной продукции и повышение безопасности производственного процесса. Охрана труда, будучи системой специальных знаний, предназначена в качестве средства обеспечения безопасности технологических процессов и производства [1]. Совершенствованию системы управления охраной труда различных предприятий постоянно уделяется внимание ученых Украины [3-6]. Одним из основных способов оценки техногенного риска на предприятиях повышенной опасности, в частности, угольных, является исследование уровня возможного травматизма.

Методика эксперимента. Химическая и горная промышленность, военные науки, планирование сколько-нибудь серьезных процессов и явлений - везде используются методы математической статистики высокого уровня, базирующиеся на использовании теории распределений, и, в первую очередь, нормальных и квазинормальных распределений. Однако становится очевидным, что по мере усложнения характера протекания рассматриваемых процессов, т. е. при вынужденном использовании других несимметричных распределений, в частности Стьюдента или Колмогорова, для описания массивов экспериментальных данных или их выборок традиционных методов статистики становится уже недостаточно и хотя непосредственно, даже при полном их соответствии к решению ряда задач горной электротехники [2], то использование их для прогноза состояния эргатических систем, например, предприятий угольной промышленности, они не позволяют дать корректную оценку рассматриваемому явлению и выполнить прогноз поведения системы в условиях, отличных от рассматриваемых.

Для того чтобы правильно понять идеи непараметрической статистики (термин был впервые введен Wolfowitz, 1942), следует познакомиться с идеями параметрической статистики. Поэтому изначально следует ознакомиться с понятием статистической значимости критерия, основанного на распределении определенной статистики (вы можете просмотреть эту главу, прежде чем продолжить чте-

ние). Говоря кратко, если вы знаете распределение наблюдаемой переменной, то можете предсказать, как в повторных выборках равного объема будет "вести себя" используемая статистика - т.е. каким образом она будет распределена. Пусть, например, имеется 100 случайных выборок, из одной популяции по 100 взрослых человек в каждой. Вычислим средний рост (возраст или стаж) субъектов в каждой выборке, т.е. построим выборочное среднее. Тогда распределение выборочных средних можно хорошо аппроксимировать нормальным распределением (более точно, t -распределением Стьюдента с 99 степенями свободы). Теперь представьте, что случайным образом извлечена еще одна выборка из жителей некоего города, где, по вашим представлениям, проживают люди с ростом выше среднего. Если средний рост людей в этой выборке попадает в верхнюю 95% критическую область t распределения, то можно сделать обоснованный вывод, что жители города, действительно, в среднем более высокие (чем в целом в популяции), т.е. что это действительно город высоких людей.

Обсуждение методики обработки объемных данных характерных для угольной промышленности Украины. Возникает вопрос: «Действительно ли большинство переменных имеют нормальное распределение? В рассмотренном примере использовался тот факт, что в повторных выборках равного объема средние значения (роста людей) будут иметь t -распределение (с определенным средним и дисперсией). Однако, это верно лишь, если рассматриваемая переменная (рост) имеет нормальное распределение, т.е. что распределение людей определенного роста нормально распределено (рис. 1).

Одним из факторов, ограничивающих применения критериев, основанных на предположении нормальности, является объем выборки. До тех пор пока выборка достаточно большая (например, 100 или больше наблюдений), можно считать, что выборочное распределение нормально, даже если вы не уверены, что распределение переменной в популяции является нормальным. Тем не менее, если выборка мала, эти критерии следует использовать только при наличии уверенности, что переменная действительно имеет нормальное распределение. Однако нет способа проверить это предположение на малой выборке.

© Б. Б. Кобылянский . 2015

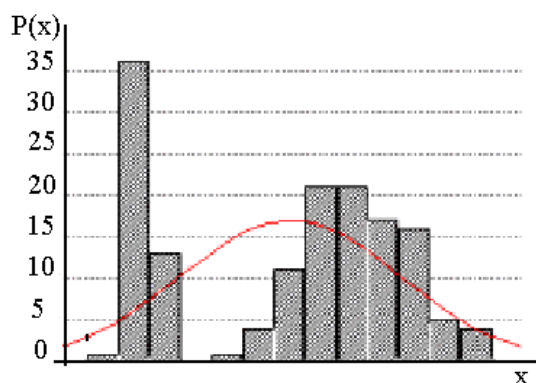


Рис. 1 – Сопоставление законов распределения

Использование критериев, основанных на предположении нормальности, кроме того, ограничено шкалой измерений. Такие статистические методы, как t -критерий, регрессия и т. д. предполагают, что исходные данные непрерывны. Однако имеются ситуации, когда данные, скорее, просто ранжированы (измерены в порядковой шкале), чем измерены точно.

Типичный пример дает группы данных: первую позицию занимает группа с максимальным числом рабочих конкретной специальности, вторую позицию занимает группа с максимальным числом рабочих среди оставшихся групп (среди групп, из которых удалена первая группа) и т. д. Зная рейтинги, а рабочих одной из групп больше числа рабочих другой, но насколько больше, сказать уже нельзя. Представьте, вы имеете 5 групп: А, В, С, D, E, которые располагаются на 5 первых мест. Пусть в текущем месяце мы имели следующую расстановку: А, В, С, D, E, а в предыдущем месяце: D, E, А, В, С. Спрашивается, произошли существенные изменения в рейтингах групп или нет? В данной ситуации, очевидно, мы не можем использовать t -критерий, чтобы сравнить эти две группы данных, и выходим в область специфических вероятностных вычислений (а любой статистический критерий содержит в себе вероятностную калькуляцию!). Мы рассуждаем примерно следующим образом: насколько велика вероятность того, что отличие в двух расстановках групп вызвано чисто случайными причинами или это отличие слишком велико и не может быть объяснено за счет чистой случайности. В этих рассуждениях мы используем лишь ранги или перестановки групп и никак не используем конкретный вид распределения числа посетителей на них.

Для анализа малых выборок и для данных, измеренных в бедных шкалах, применяют непараметрические методы, и в частности, краткий обзор непараметрических процедур. По существу, для каждого параметрического критерия имеется, по крайней мере, одна непараметрическая альтернатива. В общем, эти процедуры попадают в одну из следующих категорий: критерии различия для независимых выборок; критерии различия для зависимых выборок; оценка степени зависимости между переменными.

Вообще, подход к статистическим критериям в анализе данных должен быть прагматическим и не отягощен лишними теоретическими рассуждениями.

Имея в своем распоряжении компьютер с системой STATISTICA, вы легко примените к своим данным несколько критериев. Зная о некоторых подводных камнях методов, вы путем простого экспериментирования выберете верное решение. Развитие сюжета довольно естественно: если нужно сравнить значения двух переменных, то вы используете t -критерий. Однако следует помнить, что он основан на предположении нормальности и равенстве дисперсий в каждой группе. Освобождение от этих предположений приводит к непараметрическим тестам, которые особенно полезны для малых выборок.

Далее имеются две ситуации, связанные с исходными данными: зависимые и независимые выборки, в которых применяется t -критерий для зависимых и независимых выборок соответственно.

Развитие t -критерия приводит к дисперсионному анализу, который используется, когда число сравниваемых групп больше двух. Соответствующее развитие непараметрических процедур приводит к непараметрическому дисперсионному анализу, правда, существенно более бедному, чем классический дисперсионный анализ.

Для оценки зависимости, или, выражаясь несколько высокопарно, степени тесноты связи, вычисляют коэффициент корреляции Пирсона. Строго говоря, его применение имеет ограничения, связанные, например, с типом шкалы, в которой измерены данные, и нелинейностью зависимости, поэтому в качестве альтернативы используются также непараметрические, или так называемые ранговые, коэффициенты корреляции, применяемые, например, для ранжированных данных. Если данные измерены в номинальной шкале, то их естественно представлять в таблицах сопряженности, в которых используется критерий хи-квадрат Пирсона с различными вариациями и поправками на точность. Итак, по существу имеется всего несколько типов критериев и процедур, которые нужно знать и уметь использовать в зависимости от специфики данных. Вам нужно определить, какой критерий следует применять в конкретной ситуации. Непараметрические методы наиболее приемлемы, когда объем выборок мал. Если данных много (например, $n > 100$), часто не имеет смысла использовать непараметрическую статистику. Если размер выборки очень мал (например, $n = 10$ или меньше), то уровни значимости для тех непараметрических критериев, которые используют нормальное приближение, можно рассматривать только как грубые оценки.

Различия между независимыми группами. Если имеются две выборки (например, мужчины и женщины), которые нужно сравнить относительно некоторого среднего значения, например, среднего давления или количества лейкоцитов в крови, то можно использовать t -тест для независимых выборок. Непараметрическими альтернативами этому тесту являются критерий серий Вальда—Вольфовица, Манна—Уитни [F -тест и двухвыборочный критерий Колмогорова - Смирнова.

Различия между зависимыми группами. Если вы хотите сравнить две переменные, относящиеся к одной и той же выборке, например, медицинские показатели одних и тех же пациентов до и после приема

лекарства, то обычно используется t-критерий для зависимых выборок. Альтернативными непараметрическими тестами являются критерий знаков и критерий Вилкоксона.

Если рассматривается более двух переменных, относящихся к одной и той же выборке, то обычно используется дисперсионный анализ (ANOVA) с повторными измерениями.

Для того чтобы оценить зависимость между двумя переменными, обычно вычисляют коэффициент корреляции Пирсона. Непараметрическими аналогами коэффициента корреляции Пирсона являются коэффициенты ранговой корреляции Спирмена R, статистика Кендалла и коэффициент Гамма (более подробно см. например, книгу Кендалл М. Дж., Ранговые корреляции 1975).

Коэффициент ранговой корреляции (rank correlation coefficients) оценивает величину зависимости между переменными, измеренными в порядковых шкалах, т. е. между порядковыми переменными.

Прозрачный способ построения парных коэффициентов корреляции из обобщенного коэффициента корреляции предложил Daniels (Daniels H. E., 1948, *Biometrika*, v. 35, p. 416-417).

Обобщенный коэффициент корреляции определяется формулой:

$$\Gamma = \frac{\sum a_{ij}b_{ij}}{\sqrt{(\sum a_{ij}^2)(\sum b_{ij}^2)}} \quad (1)$$

где $a_{ij} = a(X_i, X_j)$, $b_{ij} = b(Y_i, Y_j)$ — некоторые функции пар наблюдений X и Y соответственно, суммирование ведется по всем парам i, j.

Заметим, что при $a_{ij} = X_j - X_i$, $b_{ij} = Y_j - Y_i$ получаем обычный коэффициент корреляции Пирсона. Если переменные ранжированы, то мы работаем с рангами. Упорядочим значения X_i по возрастанию, то есть построим вариационный ряд этих величин. Номер величины X_i в этом ряду называется ее рангом и обозначается R_i .

Затем упорядочим значения Y_i в порядке возрастания. Номер величины Y_{ij} в этом ряду называется ее рангом и обозначается S_i .

Коэффициент ранговой корреляции Спирмена вычисляется как обобщенный коэффициент парной корреляции с заменой наблюдений их рангами. Формально для обобщенного коэффициента корреляции нужно положить $a_{ij} = R_j - R_i$, $b_{ij} = S_j - S_i$.

Коэффициент Кендалла вычисляется, если в формуле для обобщенного коэффициента положить $a_{ij} = 1$ при $R_i < R_j$ и $a_{ij} = -1$ при $R_i > R_j$. Величины b_{ij} задаются аналогичными соотношениями с заменой рангов R_{ij} ранги S_i наблюдений Y. Итак, мы ясно видим, что идея всех корреляций возникает из одного и того же источника.

Если имеется более двух переменных, то используют коэффициент конкордации Кендалла. Например, он применяется для оценки согласованности мнений независимых экспертов (судей), например, баллов, выставленных одному и тому же участнику конкурса.

Если имеются две категориальные переменные, то для оценки степени зависимости используют стан-

дартные статистики и соответствующие критерии для таблиц сопряженности: хи-квадрат, фи-коэффициент, точный критерий Фишера.

Нелегко дать простой и однозначный совет, касающийся использования этих процедур. Каждая имеет свои достоинства и свои недостатки. Например, двухвыборочный критерий Колмогорова—Смирнова чувствителен не только к различию в положении двух распределений, но также и к форме распределения. Фактически он чувствителен к любому отклонению от гипотезы однородности, но не указывает, с каким именно отклонением мы имеем дело.

В общем, если результат исследования является важным и наблюдений немного (например, отвечает на вопрос — оказывает ли людям помощь определенная очень дорогая и болезненная лекарственная терапия?), то всегда целесообразно испытать непараметрические тесты. Возможно, результаты тестирования (разными тестами) будут различны. В таком случае следует попытаться понять, почему разные тесты дали разные результаты. С другой стороны, непараметрические тесты имеют меньшую мощность, чем их параметрические конкуренты, и если важно обнаружить даже слабые эффекты (например, при выяснении, является ли данная пищевая добавка опасной для здоровья), следует провести многократные испытания и особенно внимательно выбирать статистику критерия.

Классическая статистика хи-квадрат Пирсона замечательна тем, что ее распределение приближается распределением хи-квадрат, для которого имеются подробные таблицы. Процентные точки распределения хи-квадрат могут быть также эффективно вычислены в системе STATISTICA с помощью вероятностного калькулятора.

Свойство критерия хи-квадрат (точность аппроксимации распределения статистики распределением хи-квадрат) для таблиц 2×2 с малыми ожидаемыми частотами может быть улучшено за счет уменьшения абсолютного значения разностей между ожидаемыми и наблюдаемыми частотами на величину 0,5 перед возведением в квадрат. Это так называемая поправка Йетса на непрерывность для таблиц частот 2×2 , которая обычно применяется, когда ячейки содержат только малые частоты и некоторые ожидаемые частоты становятся меньше 5 (или даже меньше 10).

Если сумма частот небольшая, то лучше использовать точный критерий Фишера вместо критерия хи-квадрат.

Известны рекомендации Кокрена для таблиц 2×2 : если сумма всех частот в таблице меньше 20, то следует использовать точный критерий Фишера. Если сумма частот больше 40, то можно применять критерий хи-квадрат с поправкой на непрерывность. Однако эти рекомендации не универсальны [7]. Так как в данных обычно имеются ячейки с малыми частотами (2 и 3), то для улучшения точности критерия хи-квадрат используем поправку Йетса. Поскольку нас интересует односторонняя альтернатива, мы делим уровень $p = 0,0012$ пополам и получаем 0,0006.

Нелегко дать простой совет, касающийся использования непараметрических процедур. Каждая непараметрическая процедура в модуле имеет свои достоинства и свои недостатки. Например, двухвыбо-

рочный критерий Колмогорова-Смирнова чувствителен не только к различию в положении двух распределений, например, к различиям средних, но также чувствителен и к форме распределения. Критерий Вилкоксона парных сравнений предполагает, что можно ранжировать различия между сравниваемыми наблюдениями. Если это не так, лучше использовать критерий знаков. В общем, если результат исследования является важным (например, оказывает ли людям помощь определенная очень дорогостоящая и болезненная терапия?), то всегда целесообразно применить различные непараметрические тесты. Возможно, результаты проверки (разными тестами) будут различны. В таком случае следует попытаться понять, почему разные тесты дали разные результаты. С другой стороны, непараметрические тесты имеют меньшую статистическую мощность (менее чувствительны), чем их параметрические конкуренты, и если важно обнаружить даже слабые отклонения (например, является ли данная пищевая добавка опасной для людей), следует особенно внимательно выбирать статистику критерия. Большие массивы данных и непараметрические методы. Непараметрические методы наиболее приемлемы, когда объем выборок мал. Если данных много (например, $n > 100$), то не имеет смысла использовать непараметрические статистики. Главное здесь состоит в том, что когда выборки становятся очень большими, то выборочные средние подчиняются нормальному закону, даже если исходная переменная не является нормальной или измерена с погрешностью. Таким образом, параметрические методы, являющиеся более чувствительными (имеют большую статистическую мощность), всегда подходят для больших выборок. Большинство критериев значимости многих непараметрических статистик, описанных далее, основываются на асимптотической теории (больших выборок) поэтому соответствующие тесты часто не выполняются, если размер выборки становится слишком малым.

Выводы. Таким образом, из изложенного вытекает, что для обработки объемов выборок в сотни данных, характерных для угольной промышленности Украины, наиболее подходят методы непараметрической статистики.

Список литературы: 1. Минько, В. М. Охрана труда в машиностроении: учеб. для студ. учреждений сред. проф. образования [Текст] / В. М. Минько. – М.: Академия, 2012. – 256 с. 2. Типове положення про порядок проведення навчання і перевірки знань з питань охорони праці : НПАОП 0.00-4.12-05. –Х. : Форт, 2005. – 40 с. 3. Ступницька, Н. В. Підвищення ефективності планування заходів запобігання виробничому травматизму на підприємствах машинобудування : автореф. дис. ... канд. техн. наук : 05.26.01 «Охорона праці» [Текст] / Н. В. Ступницька. – Л., 1999. – 22 с. 4. Кружжилько, О. Є. Удосконалення комплексної оцінки стану охорони праці на підприємствах : автореф. дис. ... канд. техн. наук : 05.26.01 «Охорона праці» [Текст] / О. Є. Кружжилько. – К., 2001. – 20 с. 5. Гунченко, О. М. Вдосконалення системи управління охороною праці на машинобудівних підприємствах : автореф. дис. ... канд. техн. наук : 05.26.01 «Охорона праці» [Текст] / О. М. Гунченко. – Луганськ, 2007. – 20 с. 6. Касьянов, М. А. Проблеми стану і необхідності вдосконалення системи управління охороною праці в галузі машинобудування [Текст] / М. А. Касьянов, В. О. Медяник, О. М. Гунченко, Д. А. Вишневський // Вісник Східноукраїнського національного університету ім. В. Даля. – Луганськ : СНУ ім. В. Даля, 2008. – № 6 (124), Ч. 2. – С. 3–9. 7. Елисеєва, І. І. Общая теория статистики : учеб. для студ. вузов, обуч. по направл. и спец. "Статистика" [Текст] / И. И. Елисеєва, М. М. Юзбашев ; под. ред. И. И. Елисеєвой. – М.: Финансы и статистика, 1996. – 368 с. 8. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям [Текст] / Н. Ш. Кремер. – М.: ЮНИТИ-ДАНА, 2007. – 551 с. 9. Тумасян, А. А. Статистика промышленности: учебное пособие [Текст] / А. А. Тумасян, Л. И. Василевская. – Минск: Новое знание. – Москва: Инфра-М, 2012. – 429 с. 10. Холлендер, М. Непараметрические методы статистики [Текст] / М. Холлендер, Д. Вульф. – М.: Мир, 1983. – 518 с.

Bibliography (transliterated): 1. Minko, V. M. (2012). Ochrana truda v mashinostroenii: ucheb. dlya stud. uchrezhdenij sred. prof. obrazovaniya. Moscow: Akademiya, 256. 2. Tipove polozhennya pro porядok provedennya navchannya i perevirkі znan z pitan oxoroni praci : NPAOP 0.00-4.12-05. –X. : Fort, 2005. – 40 s. 3. Stupnicka, N. V. (1999). Pidvishchennya efektyvnosti planuvannya zakhodiv zapobigannya vrobnychomu travmatizmu na pidpriemstvax mashinobuduvannya: avtoref. dis. ... kand. tehn. nauk : 05.26.01 «Oxorona praci». DU «Lvivska Politehnika». Lvov, 22. 4. Kruzhylo, O. E. (2001). Udoshkonalennya kompleksnoi ocinki stanu oxoroni praci na pidpriemstvax : avtoref. dis. ... kand. tehn. nauk : 05.26.01 «Oxorona praci». Kyev, 20. 5. Gunchenko, O. M. (2007). Vdoshkonalennya sistemi upravlinnya oxoronoyu praci na mashinobudivnykh pidpriemstvax : avtoref. dis. ... kand. tehn. nauk : 05.26.01 «Oxorona praci». Lugansk, 20. 6. Kasyanov, M. A., Medyanik, V. O., Gunchenko, O. M., Vishnevskij, D. A. (2008). Problemi stanu i neobxidnosti vdoshkonalennya sistemi upravlinnya oxoronoyu praci v galuzi mashinobuduvannya. Visnik Sxidnoukrajnського nacionalnogo universitetu im. V. Dalya. Lugansk : SNU im. V. Dalya, № 6 (124), ch. 2, 3–9. 7. Eliseeva, I. I., Yuzbashev, M. M. (1996). Obshhaya teoriya statistiki : ucheb. dlya stud. vuzov, obuch. po napravl. i spec. "Statistika. Moscow: Finansy i statistika, 368. 8. Kremer, N. Sh. (2007). Probability theory and mathematical statistics: a textbook for university students enrolled on economic specialties. Moscow: UNITY-DANA, 551. 9. Tumasyan, A., Wasilewska, L. I. (2012). Industrial statistics: a manual. Minsk: The new knowledge. Moscow: Infra-M, 429. 10. Hollender, M., Wolfe, D. (1983). Non-parametric statistical methods. Moscow: Mir, 518.

Поступила (received) 20.12.2015

Відомості про авторів / Сведения об авторах / About the Authors

Кобылянский Борис Борисович – Кандидат технічних наук, Учебно-научний професійно-педагогічний інститут Української інженерно-педагогічної академії, доцент кафедри охорони праці та екологічної безпеки; тел.: 050-025-29-06; e-mail: b.kobiliansky@yandex.ua

Kobilyansky Boris Borisovich, PhD tehnycheskyh Sciences, Teaching and Research Professional Pedagogical Institute Ukrainian engineering and Pedagogical Academy, assistant professor of occupational and environmental safety; Tel. : 050-025-29-06; e-mail: b.kobiliansky@yandex.ua