



Б. Б. КОБЫЛЯНСКИЙ,
канд. техн. наук
(УНППИ Украинской инженерно-педагогической академии)



А. Г. МНУХИН,
доктор техн. наук
(Запорожская государственная инженерная академия)

На современных предприятиях угольной промышленности, оснащенных автоматизированными линиями и технологическими комплексами, внедряются также гибкие производственные системы, что способствует решению двух взаимосвязанных задач: выпуску более качественной продукции и повышению безопасности производственного процесса. Ученые Украины уделяют постоянное внимание охране труда как средству обеспечения безопасности выполнения технологических процессов [1–4].

Один из основных способов оценки техногенного риска на угольных предприятиях – исследование уровня возможного травматизма. Известно, что

УДК 621.3:622:519.24

Применение методов непараметрической статистики для оценки некоторых особенностей горного производства

Для управления сложными технологическими многофакторными системами угольной промышленности впервые предлагается использование методов непараметрической статистики, позволяющих выполнить корректный прогноз исследуемых явлений за пределы имеющихся экспериментальных данных с достаточной для решения большинства практических задач точностью, особенно в тех случаях, когда количество данных относительно невелико.

Ключевые слова: угольная промышленность, оценка особенностей горного производства, непараметрическая статистика.

Контактная информация: anatoly.mnukhin@gmail.com

высокой аварийности и травматизму на шахтах страны, особенно Донбасса, способствуют сложнейшие горно-геологические условия разработки угольных пластов, устаревший шахтный фонд и горношахтное оборудование, а также другие факторы, которые следует учитывать при анализе травматизма. Кроме того, изучение одного или группы факторов требует разных статистических подходов и математических методов для их практической реализации.

В угольной промышленности большинство предприятий классифицируют как шахты с особо опасными условиями труда: 89 % – газовые, 60 % – опасные по угольной пыли, 45,2 % – опасные по газодинамическим явлениям (внезапные выбросы угля, породы и газа, горные удары) и 22,6 % – по самовозгоранию угля (табл. 1).

Анализ травматизма и аварийности свидетельствует о значительном удельном весе происшествий, произошедших по причине человеческого фактора, особенно аварий с групповым травматизмом. Это подтверждают материалы расследования аварий I и II категорий с групповыми несчастными случаями, согласно которым одна из причин аварий и травматизма – нарушение правил техники безопасности. Для установления основных причин и тенденции изменения этих явлений, а также их группировки необходимо глубокое изучение условий, в которых традиционные методы элементарной статистики недостаточно эффективны.

Очевидно, что по мере усложнения характера протекания рассматриваемых процессов, т. е. при вынужденном использовании несимметричных распределений, в частности Стьюдента или Колмогорова, для описания массивов экспериментальных данных или их выборок традиционных методов статистики уже недостаточно. Даже при полном их соответствии решению ряда за-

Таблица 1

Регион	Общее количество шахт с особо опасными условиями труда	Количество шахт – технических единиц по видам опасности								
		Выделение метана					Внезапные выбросы угля, газа и породы	Горные удары	Взрывчатая угольная пыль	Самовозгорание угля
		Всего шахт	В том числе по категориям							
	I		II	III	Сверхкатегорийные					
Донецкая область	105	47	10	5	3	29	53	13	65	27
Луганская область	70	34	5	2	4	23	20	–	34	11
Другие области	15	15	4	4	4	3	–	–	15	5
Украина в целом	190	96	19	11	11	55	73	13	114	43
Удельный вес шахт, %	100,0	50,5	10,0	5,8	5,8	28,9	38,4	6,8	60,0	22,6

дач горной электротехники [5] использование методов математической статистики, а именно теории распределений, для прогноза состояния эргатических систем, например предприятий угольной промышленности, не позволяет получить корректную оценку рассматриваемого явления и выполнить прогноз поведения системы в условиях, отличных от рассматриваемых.

В мире ежегодно травмируется до 200 тыс. горняков, в том числе 6–10 тыс. – со смертельным исходом. Наибольшее количество аварий происходит на угольных шахтах Китайской Народной Республики, где ежедневно в результате несчастных случаев гибнет более 10 шахтеров, а в 1995 г. число погибших превысило 10 тыс. 400 человек.

За последние 30 лет даже на относительно малоаварийных польских шахтах произошло 15 взрывов метановоздушной смеси, при этом пострадало 426 человек, в том числе 219 – смертельно [6]. В этот период наблюдалось 17 катастрофических взрывов газа и пыли на шахтах Российской Федерации, 4 – в Югославии, 5 – в Японии, 3 – в ЮАР, а также в ФРГ, Турции, Перу, Тайване, Мозамбике [6].

Вместе с тем в ведущих угледобывающих странах в последние годы количество аварий и катастроф значительно снизилось. В табл. 2 приведены данные по среднегодовому (2000–2007 гг.) количеству аварий и простоев на шахтах Украины. Для анализа этих данных, харак-

Таблица 2

Причины аварий и простоев	Количество случаев		Потери	
	Всего	%	рабочего времени, ч	добычи, тыс. т
Аварии и простои эксплуатационные	29 440	26,2	208 859	3 745
В том числе завалы (обрушения) в лавах	6 394	–	43 803	832
Отказы горношахтного оборудования	34 800	31,0	178 757	4 023
В том числе забойного оборудования	28 062	–	145 436	3 339
Аварии и простои на главных подъемных установках	548	0,4	5 114	111
Аварии (простои) на шахтном транспорте	19 005	16,9	97 028	2 471
В том числе на конвейерном транспорте	12 871	–	59 006	1 934
Остановки по предписанию инспекции, участка вентиляции и техники безопасности (ВТБ)	4 162	3,8	44 113	969
Прочие опасности, всего	24 314	21,7	208 887	3 335
В том числе необеспеченность крепежным лесом	1 817	–	9 384	212
Остановки вентиляторов главного проветривания	211	–	7 033	42
Загазованность выработок	143	–	2 321	24
Затопление (подтопление) выработок	432	–	6 991	197

теризуемых малым объемом, требуются иные математические методы, в частности методы непараметрической статистики.

Термин «непараметрическая статистика» впервые был введен И. Вольфовицем в 1942 г. [7]. Изначально ознакомимся с идеями *параметрической* статистики, в частности с понятием «статистическая значимость критерия», основанном на распределении определенной статистики. Так, если известно распределение наблюдаемой переменной, то можно 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 одного множества по 100 человек в каждой. Вычислим среднюю квалификацию (возраст или стаж) субъектов в каждой выборке, т. е. построим выборочное среднее. Тогда распределение выборочных средних можно аппроксимировать нормальным распределением (более точно, *t*-распределением Стьюдента с 99 степенями свободы).

Допустим, что случайно извлечена еще одна выборка из числа трудящихся предприятия, где, как кажется, работают люди возрастом выше среднего. Если средний возраст людей в этой выборке попадает в верхнюю (95 %) критическую область *t*-распределения, то можно сделать вывод, что работники этого объединения действительно имеют в среднем более высокую квалификацию, чем в целом в промышленности, т. е. что это действительно наиболее хорошо функционирующее предприятие.

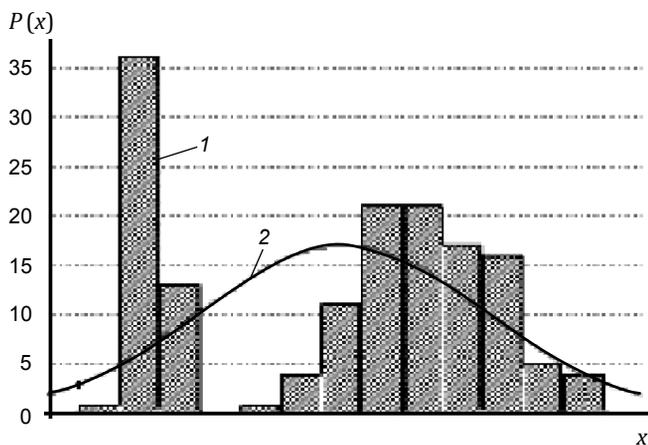


Рис. 1. Сопоставление законов распределения: 1 и 2 – экспериментальные и расчетные значения.

Возникает вопрос: действительно ли большинство переменных имеют нормальное распределение. В рассмотренном примере использовался факт, что в повторных выборках равного объема средние значения возраста людей будут иметь *t*-распределение (с определенным средним и дисперсией). Однако это верно лишь в случае, если переменная имеет нормальное распределение, т. е. если распределение людей определенной квалификации можно описать нормальным законом (рис. 1).

Одним из факторов, ограничивающих применение критериев, основанных на предположении нормальности, является объем выборки. До тех пор, пока выборка достаточно большая (например, 100 наблюдений и более), можно считать, что выборочное распределение нормально, даже если нет уверенности, что распределение переменной группы работников нормально. Если же выборка мала, то эти критерии следует использовать только при наличии уверенности, что переменная действительно имеет нормальное распределение. Однако нет надежного способа проверить это предположение на малой выборке.

Использование критериев, основанных на предположении нормальности, кроме того, ограничено шкалой измерений. Такие статистические методы, как *t*-критерий, регрессия и другие предполагают, что исходные данные непрерывны. Однако есть ситуации, когда данные просто ранжированы (измерены в порядковой шкале), а не измерены точно.

Типичный пример дают группы данных: первую позицию занимает группа с максимальным количеством рабочих конкретной специальности, вторую – группа с максимальным количеством рабочих среди оставшихся групп, т. е. групп, из которых удалена первая, и т. д. Из этих данных узнаем рейтинги, что число рабочих одной из групп больше числа рабочих другой, но насколько больше, определить нельзя.

Рассмотрим пять групп: *A, B, C, D* и *E*, которые располагаются на пяти первых местах. Пусть в текущем месяце была следующая расстановка: *A, B, C, D, E*, а в предыдущем – *D, E, A, B, C*. Необходимо установить – произошли существенные изменения в рейтингах групп или нет. В данной ситуации, очевидно, что нельзя использовать *t*-критерий, чтобы сравнить эти две группы данных, поэтому надо перейти в область специфических вероятностных вычислений.

Для анализа малых выборок и данных, измеренных в «бедных» шкалах, применяют непараметрические методы, т. е. используют непараметрические процедуры. По существу, для каждого параметрического критерия имеется, по крайней мере одна, непараметрическая альтернатива.

Непараметрические методы наиболее приемлемы, когда объем выборок мал. Если данных достаточно (например, $n > 100$), часто не имеет смысла использовать непараметрическую статистику. Если размер выборки очень мал (например, $n = 10$ и менее), то уровни значимости для тех непараметрических критериев, которые используют нормальное распределение, можно рассматривать только как грубые оценки.

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.

Аварии на угольных шахтах относятся к случайным событиям. Поэтому, чтобы иметь возможность объективно сравнивать отдельные шахты, группы шахт (объединений, холдинговых компаний) или изучать отрасль в целом по аварийности, необходимо определить статистически подтвержденные показатели аварий-

ности. Изучив соответствующие научные источники (например, [8]), можно сделать вывод о необходимости использования для анализа данных о горном производстве разных статистических распределений – как нормального (логарифмически нормального, экспоненциального, Вейбулла, гамма-распределения), так и других, в том числе Стьюдента, Пирсона.

Для проверки высказанных положений данные по аварийности в горной промышленности Украины по 60 крупным шахтам обрабатывались посредством компьютерных технологий с использованием разных законов распределения вероятности аварий. Результаты расчетов приведены в табл. 3, из которой следует, что дальнейшая обработка данных и последующий анализ могут быть реализованы только методами непараметрической статистики.

Отличительная особенность классической статистики хи-квадрат Пирсона заключается в том, что ее распределение приближается к распределению хи-квадрат, для которого имеются подробные таблицы. Процентные точки распределения хи-квадрат можно также эффективно вычислить в системе STATISTICA.

Свойство критерия хи-квадрат (точность аппроксимации распределения статистики распределением хи-квадрат) для таблиц 2×2 с малыми ожидаемыми частотами можно улучшить путем уменьшения абсолютного значения

Таблица 3

Закон распределения вероятности аварий	Параметры распределения		Оценка достоверности				Уровень значимости	Оценка соответствия
	Наименование	Значение	Число степеней свободы	Критерий Пирсона				
				фактический	теоретический при уровне значимости			
		0,05	0,01					
Нормальный	Среднее	413,707	10	148,651	18,3	23,2	0	Не соотв.
	Стандартное отклонение	-	-	-	-	-	-	-
Логарифмически нормальный	Среднее	451,614	12	24,762	21,0	26,2	$7,30 \cdot 10^{-3}$	Не соотв.
	Стандартное отклонение	541,425	-	-	-	-	-	-
Экспоненциальный	Среднее	413,700	13	45,670	22,4	27,7	$1,62 \cdot 10^{-3}$	Не соотв.
Вейбулла	Форма (α)	1,267	10	16,782	18,3	23,2	0,073	Соотв.
	Масштаб (β)	$446,861 \cdot 10^{-3}$	-	-	-	-	-	-
Гамма-распределение	Форма (α)	1,415	10	15,62	18,3	23,2	0,111	Соотв.
	Масштаб (β)	$8,420 \cdot 10^{-3}$	-	-	-	-	-	-

разностей между ожидаемыми и наблюдаемыми частотами на значение 0,5 перед возведением в квадрат. Эта непрерывность применяется, когда ячейки содержат только малые частоты и некоторые ожидаемые частоты становятся меньше 5 (или меньше 10).

Согласно рекомендациям Кохрена для таблиц 2×2 : если сумма всех частот в таблице меньше 20, то вместо критерия хи-квадрат следует использовать точный критерий Фишера; если же сумма частот больше 40, то можно применять критерий хи-квадрат с поправкой на непрерывность.

Поскольку в данных обычно есть ячейки с малыми частотами (2 и 3), то для улучшения точности критерия хи-квадрат используем указанную поправку. Так как нас интересует односторонняя альтернатива, делим уровень $p = 0,0012$ пополам и получаем 0,0006.

Однозначное предложение по использованию непараметрических процедур дать нелегко. Каждая такая процедура в модуле имеет свои достоинства и свои недостатки. Так, двухвыборочный критерий Колмогорова–Смирнова чувствителен не только к различию в выборе одного из двух распределений, например к различиям средних, но и к форме распределения. Критерий Вилкоксона парных измерений предполагает, что можно ранжировать различия между сравниваемыми наблюдениями. Если это не так, лучше использовать критерий знаков. С одной стороны, если результат исследования имеет важное значение, то целесообразно применить разные 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, являются ли данные отклонения входных факторов опасными для людей), нужно особенно внимательно выбирать статистику критерия.

В целом непараметрические методы наиболее приемлемы, когда объем выборки относительно невелик. Если данных много (например, $n > 100$), то не имеет смысла использовать непараметрические статистики. Главное состо-

ит в том, что когда выборки становятся очень большими, то выборочные средние подчиняются нормальному закону, даже если исходная переменная не является нормальной или измерена с определенной погрешностью. Таким образом, параметрические методы как более чувствительные (имеют большую статистическую мощность) всегда подходят для больших выборок. Критерии значимости многих непараметрических статистик основываются на асимптотической теории (больших выборок), поэтому соответствующие тесты часто не выполняются, если размер выборки становится слишком мал.

Выводы. Для обработки объемов нестабильных выборок даже в сотни данных, характерных для угольной промышленности Украины, наиболее подходят методы непараметрической статистики.

Л И Т Е Р А Т У Р А

1. *Ступницька Н. В.* Підвищення ефективності планування заходів запобігання виробничому травматизму на підприємствах машинобудування: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.26.01 «Охорона праці» / Н. В. Ступницька. – Львів, 1999. – 22 с.
2. *Кружилко О. Є.* Удосконалення комплексної оцінки стану охорони праці на підприємствах: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.26.01 «Охорона праці» / О. Є. Кружилко. – К., 2001. – 20 с.
3. *Гунченко О. М.* Вдосконалення системи управління охороною праці на машинобудівних підприємствах: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.26.01 «Охорона праці» / О. М. Гунченко. – Луганськ, 2007. – 20 с.
4. *Касьянов М. А.* Проблеми стану і необхідності вдосконалення системи управління охороною праці в галузі машинобудування / М. А. Касьянов, В. О. Медяник, О. М. Гунченко, Д. А. Вишнівський // Вісн. Східноукр. нац. ун-ту ім. В. Даля. – Луганськ: СНУ ім. В. Даля, 2008. – Ч. 2, № 6 (124). – С. 3–9.
5. *Типове положення про порядок проведення навчання і перевірки знань з питань охорони праці:* НПАОП 0.00-4.12-05. –Х.: Форт, 2005. – 40 с.
6. *Урбанчик Ю.* Мала механізація в горнодобиваючій промисловості в світлі техніки безпеки і продуктивності праці: пер. с пол. / Ю. Урбанчик. – Катовице: Механізація, 2001. – С. 7–9.
7. *Холлендер М.* Непараметрические методы статистики / М. Холлендер, Д. А. Вулф. – М.: Финансы и статистика, 1983. – 518 с.
8. *Левкин Н. Е.* Предотвращение аварий и травматизма в угольных шахтах Украины / Н. Е. Левкин. – Макеевка: МакНИИ, 2002. – 392 с.