

Культура статистического анализа и безопасности

В.А. Машин

Машин Владимир Анатольевич - Главный специалист Центрального Института Повышения Квалификации Госкорпорации «РОСАТОМ». Кандидат психологических наук.

E-mail: mashin-va@mail.ru

Введение

Некорректное использование методов статистического анализа в научных исследованиях представляет собой давнюю и широко распространенную проблему (Леонов В.П., Ижевский П.В., 1997; Гланц С., 1998; Леонов В.П., 2002а; Ланг Т., 2005). Одна из ее причин это низкая статистическая культура (Леонов В.П., 2002b), отражающая степень, в которой исследователи следуют принципам математической статистики и требованиям в области планирования исследований, обработки данных и представления результатов статистического анализа (Сидоренко Е.В., 2003; Бащинский С.Е., Власов В.В., 2006; Ланг Т.А., Сесик М., 2011). Последовательное и осознанное выполнение этих требований предполагает наличие специфических знаний и умений для применения на практике методов статистики, включая выбор статистических критериев на основе объема репрезентативной выборки, характера исследуемого признака (количественный, порядковый или качественный) и типа распределения (нормальное или нет) (Гласс Дж., Стэнли Дж., 1976). Эти базовые знания и умения также необходимы, чтобы понимать и оценивать правильность применения статистических методов, используемых для анализа результатов, представленных в материалах исследований, и обоснования конечных выводов (Гланц С., 1998). Без этого исследователь или читатель будет испытывать лишь мистический трепет перед математическим аппаратом, лежащим в основе статистических расчетов, полагая, что точная наука не может ошибаться, и наивно веря во всемогущество статистических методов без понимания их роли и особенностей применения (Бащинский С.Е., 1998).

В данной статье мы рассмотрим вопросы корректности применения статистических методов на примере процедуры валидации и оценки надежности опросника, который был разработан Институтом по эксплуатации АЭС США (Institute of Nuclear Power Operations - INPO) для оценки культуры безопасности. Независимый отраслевой институт INPO был создан после аварии на АЭС "Три-Майл Айленд" (США) в 1979 года, которая привела к частичному расплавлению активной зоны реактора. INPO занимается формулированием соответствующих стандартов безопасности для управления, обеспечения качества,

эксплуатационных процедур и практик, а также для проведения независимых оценок объектов атомной энергетики. Кроме этого INPO систематически собирает, рассматривает, анализирует и распространяет информацию по инцидентам, которые происходят в атомной отрасли, а также по положительным практикам.

Прежде чем мы перейдем к рассмотрению применения статистического анализа к опроснику культуры безопасности INPO, кратко остановимся на самой концепции культуры безопасности.

Понятие культуры безопасности

Термин "культура безопасности" был впервые использован в 1986 году в «Итоговом докладе о совещании по рассмотрению причин и последствий аварии в Чернобыле» (IAEA, 1986), подготовленным Международной консультативной группой по ядерной безопасности (International Nuclear Safety Advisory Group - INSAG), созданной при Генеральном директоре МАГАТЭ (Международное агентство по атомной энергии или International Atomic Energy Agency - IAEA). В 1991 году в докладе INSAG «Культура безопасности» (IAEA, 1991) было дано определение этому термину, представлена концепция Культуры безопасности, ее универсальные черты, основные компоненты и требования к ним, а также обширный перечень вопросов для оценки эффективности культуры безопасности. Культура безопасности была определена как "такой набор характеристик и аттитюдов деятельности организаций и поведения отдельных лиц, который устанавливает, что проблемам безопасности атомных станций, как обладающим высшим приоритетом, уделяется внимание, определяемое их значимостью".

В глоссарии к Руководству МАГАТЭ "Подготовка персонала атомных станций и его оценка" (IAEA, 1996) было приведено определение термина "аттитюд". В контексте культуры безопасности оно звучит следующим образом: "аттитюды это личные чувства, ощущения, представления, ценности и интересы человека, которые позволяют ему выполнять предписанные работу или задачу, уделяя внимание проблемам безопасности, как обладающим высшим приоритетом и значимостью". Набор характеристик, который фигурирует в определении культуры безопасности, включает в себя (IAEA, 1991): "преданность делу", "направленное на безопасность мышление", "стремление к совершенству" и "чувство персональной ответственности".

Отдавая себе отчет в "неосвязаемости" перечисленных характеристик и аттитюдов, INSAG посвятила основную часть своего доклада (IAEA, 1991) описанию их "ощутимых проявлений" и попыткам наполнить реальным содержанием показатели культуры безопасности: было предложено 18 показателей и 143 вопроса для их оценки. В руководстве МАГАТЭ 1994 года по самостоятельной оценке культуры безопасности в организациях и в

рамках специальных миссий агентства (МАГАТЭ, 1994) эти вопросы были дополнены 296 "наводящими вопросами" и 378 "ключевыми индикаторами" для количественной оценки характеристик и аттитюдов культуры безопасности.

Достижениям и положительной практике в области самооценки культуры безопасности был посвящен отчет, подготовленный МАГАТЭ в 2002 году (МАГАТЭ, 2002). Среди методов, которые были рекомендованы для этой цели, значились и опросники. Казалось, процесс развития и совершенствования культуры безопасности на атомных станциях, с использованием эффективных методов оценки, приобрел необратимый характер.

Но в марте 2002 года на крышке реактора АЭС Дэйвис-Бессии (США) персоналом станции была обнаружена раковина, коррозионного происхождения (ВАО АЭС, 2003). В этом месте углеродистая сталь крышки реактора, толщиной около 15 см, была полностью разрушена. Оставался лишь один слабый барьер для теплоносителя, находящегося под высоким давлением (173.4 кгс/см²) - слой нержавеющей стали толщиной менее 1 см с внутренней стороны корпуса реактора. Как было установлено, приблизительно в течение 6 лет небольшая течь теплоносителя, содержащего борную кислоту, разъедала крышку реактора. Станция, фактически, была в шаге от крупнейшей после Чернобыля аварии. Потеря теплоносителя, выполняющего функцию охлаждения, могла привести к расплавлению активной зоны реактора.

Расследование события выявило грубые нарушения Культуры безопасности на всех уровнях управления и надзора за деятельностью АЭС Дэйвис-Бессии, включая Комиссию по ядерному регулированию США (Nuclear Regulatory Commission – NRC: независимое агентство Федерального правительства США, регулирующее работу АЭС через лицензирование, надзор и обеспечение соблюдения своих требований). Итоговые выводы заставили пересмотреть существующий подход к культуре безопасности и более ясно и четко определить те критерии, по которым она должна оцениваться и контролироваться. В результате, в 2004 году INPO выпускает документ "Принципы сильной культуры ядерной безопасности", в котором формулирует 8 принципов и 56 атрибутов сильной культуры безопасности (INPO, 2004). В 2006 году этот документ стал основой Методических указаний (WANO, 2006), опубликованных Всемирной Ассоциацией Операторов АЭС (ВАО АЭС или World Association of Nuclear Operators - WANO), в которую входят все организации мира, эксплуатирующие около 450 атомных энергоблоков. В 2006 году МАГАТЭ подготовило Руководство по безопасности "Применение систем управления для установок и деятельности" (IAEA, 2006), в котором определила 5 важнейших принципов и 38 атрибутов сильной культуры безопасности. В 2008 году МАГАТЭ в Руководстве для экспертных групп по оценке культуры безопасности (IAEA, 2008) дополнила эти принципы и атрибуты перечнем из 307 уточняющих вопросов. Наконец, NRC выработала 9 черт сильной культуры,

которые вошли в Заявление о политике в области Культуры безопасности, окончательный вариант которого был опубликован в 2011 году в Федеральном Реестре США (NRC, 2011).

В 2006 году ENEL (итальянская энергетическая компания, одна из крупнейших в мире) совместно с парижским центром ВАО АЭС разработала опросник для оценки ядерной культуры безопасности, который основывался на принципах и атрибутах, предложенных в 2004 году INPO (SCK/CEN, 2012). Наряду с атомной отраслью, многочисленные опросники культуры безопасности были созданы в таких областях человеческой деятельности, как авиация, здравоохранение, газовая и нефтяная промышленность, железнодорожный транспорт (Keil Centre, 2003; HSE, 2005; Singla A.K., 2006; EUNETPAS, 2010; Health Foundation, 2011; Fricke-Ernst C., Kluge A., 2012; SCK/CEN, 2012; SNL, 2013; NLR, 2013).

Однако анализ существующих опросников культуры безопасности (SCK/CEN, 2012) выявил отсутствие для подавляющего большинства информации относительно критериального типу валидности: соответствия результатов оценки культуры безопасности определенным значениям критериальной переменной или вероятности критериального события. Разработчики опросников зачастую обходят своим вниманием прогностическую валидность – способность по результатам оценки культуры безопасности прогнозировать критериальное событие (например, судить о показателях безопасности производства спустя определенное время после измерения), текущую валидность – соответствие результатов опроса независимому критерию (например, показателям безопасной эксплуатации) в момент измерения, а также дискриминативность инструмента - способность дифференцировать организации или группы с различным уровнем безопасности (например, по количеству аварий или несчастных случаев).

Национальная лаборатория Министерства энергетики США Sandia National Laboratories (SNL) в своем недавнем обзоре исследований по вопросам культуры безопасности (SNL, 2013) пришла к неутешительным выводам: несмотря на то, что концепция культуры безопасности была предложена почти 25 лет назад (IAEA, 1991), на сегодняшний день отсутствует общая методология и методика определения и оценки уровня культуры безопасности; существует крайне мало публикаций, посвященных отношению культуры безопасности и показателей эффективности и безопасности деятельности предприятий, а имеющиеся результаты носят, в лучшем случае, неопределенный характер.

В этой связи представляет особый интерес опросник культуры безопасности, разработанный INPO в 2010 году по заказу института атомной энергетики США (Nuclear Energy Institute – NEI, главный лоббист атомной отрасли в США) в рамках исследования культуры безопасности персонала атомных станций страны. Основными целями этого исследования было изучение факторов, которые определяют понятие культуры безопасности

в атомной отрасли и степени, в которой они соответствуют 9 чертам сильной культуры безопасности, сформулированными NRC (NRC, 2011):

1. Ценность безопасности для руководителей и реальное поведение,
2. Выявление и разрешение проблем,
3. Персональная ответственность,
4. Рабочие процессы,
5. Постоянное обучение,
6. Среда, способствующая сообщению о проблемах,
7. Эффективная коммуникация в вопросах безопасности,
8. Уважительная рабочая среда,
9. Критическая позиция.

Важной задачей исследования было также оценить отношения между факторами культуры безопасности, выявленными с помощью опросника, и показателями эффективности деятельности АЭС (включая вопросы безопасности).

NRC, заинтересованная в развитии и совершенствовании культуры безопасности на контролируемых АЭС, решила выполнить независимую оценку исследованиям INPO в области культуры безопасности, включая анализ валидности и надежности разработанного INPO опросника. Для этих целей INPO любезно предоставил данные своего исследования культуры безопасности персонала АЭС США.

Результаты независимой оценки, а также описание используемых статистических методов из пакета SPSS (Statistical Package for the Social Sciences) были изложены в специальном отчете, который NRC выложила для открытого доступа на своем сайте (Morrow S., Barnes V., 2012). На вопросах корректности использования представленных в этом отчете статистических методов и обоснованности полученных выводов мы остановимся ниже.

Анализ опросника культуры безопасности

Первоначально опросник культуры безопасности INPO содержал 110 утверждений. Для ответов использовалась 7-балльная шкала Лайкерта ("совершенно не согласен", "не согласен", "скорее не согласен", "ни то, ни другое", "скорее согласен", "согласен", "совершенно согласен") и нулевой выбор: "не знаю, не имею возможности наблюдать".

Процедура опроса и сбора данных была спланирована и проведена INPO. В опросе принял участие персонал 63 станций из 103 действующих в США АЭС. Анкеты опросника были разосланы через интернет 6333 случайно выбранным сотрудникам АЭС. В ответ было получено 3031 анкета; из них отобрано 2876, которые были полностью или почти полностью заполнены по 110 пунктам. В среднем, от каждой станции, кроме двух, было получено не менее 30 анкет. После первичного анализа методом главных компонент (Principal Components

Analysis - PCA), было принято решение использовать в дальнейшем анализе лишь результаты по 60 пунктам опросника.

Объем выборки

При определении размера выборки по каждой АЭС INPO опиралось на утверждение Центральной Предельной Теоремы, что по мере того, как размер случайной выборки возрастает, выборочное среднее приближается к среднему генеральной совокупности. Когда размер случайной выборки становится больше 30, она, приблизительно, имеет нормальное распределение и ее среднее значение незначительно отличается от среднего значения генеральной совокупности. Таким образом, согласно INPO, размер выборки равный 30 может использоваться в качестве минимального порога, который устанавливается, чтобы выборка представляла собой хорошую оценку для генеральной совокупности.

Но суть в том, что Центральная Предельная Теорема не используется для определения размера выборки в исследованиях генеральной совокупности (Ивановский Р.И., 2008; Сенатов В.В., 2009). И выбор минимального порога для размера выборки $N \geq 30$ является совершенно произвольным. Формулировка Центральной Предельной Теоремы не содержит выводов относительно размера выборки $N \geq 30$, формы ее распределения и отличий среднего от среднего генеральной совокупности.

Если средние показатели и дисперсия для выборочной и генеральной совокупности не известны, для определения минимального объема выборки, обеспечивающего ее достаточную репрезентативность (соответствие характеристик выборочного наблюдения показателям, характеризующим всю генеральную совокупность) можно воспользоваться таблицами опытно полученных данных (Москвин С., 1993; Гончаров А.И., 1993; Носс И.Н., 2003). В этом случае, при средней численности персонала на АЭС США 500 человек, объем минимальной репрезентативной выборки составит 222 человека.

Другой способ, воспользоваться онлайн калькулятором для расчета размера выборки, например, компании Creative Research Systems (<http://www.surveysystem.com/sscalc.htm>), который учитывает доверительные интервал, доверительную вероятность и размер генеральной совокупности (Bland M., 2000; Lenth R.V., 2001):

- *Доверительный интервал* задает ширину диапазона по обе стороны от выбранной точки, куда могут попадать ответы. Здесь имеются в виду так называемые случайные ошибки, которые связаны с природой любых статистических погрешностей. Допускаемая погрешность производимых наблюдений (или ошибка репрезентативности), вычисляемая как доля от выборочной средней, задается в пределах от 0,01 ($\pm 1\%$) до 0,05 ($\pm 5\%$) (Бурлачук Л.Ф., Морозов С.М., 1999).

- *Доверительная вероятность* показывает, с какой вероятностью ответ попадет в границы доверительного интервала. Например, при доверительной вероятности 95%, при проведении 100 исследований в 95 из них получаемые ответы будут находиться в заданных пределах доверительного интервала. Обычно применяют 95% и 99% доверительную вероятность. Практически достаточная величина для подавляющего большинства исследований 95% (Бурлачук Л.Ф., Морозов С.М., 1999).

При объеме генеральной совокупности 500 человек, доверительном интервале 5% и доверительной вероятности 95%, размер выборки, согласно калькулятору, составит 217 человек, что очень близко к данным, полученным опытным путем. Но важно помнить, что используемые в онлайн калькуляторах формулы для расчета размера выборки основаны на допущении, что ответы респондентов близки к нормальному распределению, а это требует отдельной проверки (Малхорта Н.К., 2002).

Таким образом, выбор INPO минимального объема выборки в 30 человек с каждой АЭС не отвечает требованиям репрезентативности и не позволяет выборочным наблюдениям с приемлемой точностью отражать характеристики генеральной совокупности.

Тип данных

В ходе исследования INPO культуры безопасности, персонал АЭС просили ответить на пункты опросника, оценив степень согласия с каждым утверждением по 7-балльной шкале Лайкерта, проранжированной от выраженного несогласия до выраженного согласия. Согласно INPO, система баллов от 1 до 7 позволяет исследователю приписать количественные значения для качественных оценок респондентов. Эти данные использовались в дальнейшем анализе в качестве приближения к интервальной шкале, когда варианты ответов располагаются на относительно равном расстоянии друг от друга по шкале оценок.

Необходимо напомнить (Мещеряков Б.Г., Зинченко В.П., 2003; Bond T.G., Fox C.M., 2013), что шкала Лайкерта была предложена в 1932 году для измерения аттитюдов и представляет собой порядковую (или ранговую) шкалу, которая указывает лишь последовательность (ранг) в оценках исследуемого признака, поскольку отсутствует стандарт (эталон) для количественного измерения различий между ними:

$$R_2 - R_1 \neq R_3 - R_2 \neq R_4 - R_3 \neq R_5 - R_4 \neq R_6 - R_5 \neq R_7 - R_6.$$

Измерения в интервальной шкале не только упорядочены по рангам, но и разделены определенными интервалами (Гласс Дж., Стэнли Дж., 1976).

$$R_2 - R_1 = R_3 - R_2 = R_4 - R_3 = R_5 - R_4 = R_6 - R_5 = R_7 - R_6.$$

В интервальной шкале установлены единицы измерения, имеющие размерность (градус, секунда и т.д.). Измеряемому здесь объекту присваивается число, равное количеству единиц измерения, которое он содержит.

Порядковая шкала Лайкерта позволяет оценить выраженность аттитюдов, при этом из статистических процедур допускается, например, определение медиан распределений, вычисление коэффициентов ранговой корреляции и сопряженности (Мещеряков Б.Г., Зинченко В.П., 2003).

Таким образом, обоснование порядковой шкалы Лайкерта в качестве интервальной, с последующим применением к выборочным данным статистических критериев, разработанных для интервальных шкал, является некорректным.

Метод главных компонент

Подход INPO к определению размерности конструкта культуры безопасности, на основании полученных ответов персонала АЭС на утверждения опросника, базировался на анализе главных компонент (РСА). Но при этом не были рассмотрены условия, необходимые для использования РСА (Лоули Д.Н., Максвелл А.Э., 1967; Иберла К., 1980; Ким Дж.-О. и др., 1989; Rencher A.C., 2002):

- Проверка допущения о многомерном нормальном распределении исходных переменных;
- Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных;
- Проверка целесообразности применения факторной модели анализа к исходным данным;
- Выполнение процедуры отбора общих факторов.

Из представленного отчета неясно, какой тип корреляции был использован при анализе главных компонент. Поскольку в других местах отчета упоминается коэффициент корреляции Пирсона, можно косвенно предположить, что и при анализе главных компонент мы имеем дело с этим типом корреляции. Но использование корреляции Пирсона для порядковых данных является не корректным. Корреляция Пирсона используется для интервальных данных. Для порядковых данных необходимо использовать ранговые коэффициенты корреляции Спирмена или Кендалла.

При описании результатов анализа главных компонент по данным ответов респондентов на 110 утверждений опросника в отчете отсутствуют матрица факторных нагрузок, матрица собственных значений и матрица значений общности, что является необходимым для понимания и обоснования полученного факторного решения. Из текста

лишь следует, что на 9 первых компонент или факторов (по количеству черт сильной культуры безопасности, предложенных NRC) приходилось только 58% дисперсии данных. Это означает, что 42% дисперсии в ответах респондентов на вопросы анкеты по культуре безопасности были обусловлены другими факторами. Какое либо обоснование отбора данного количества факторов отсутствует.

Поскольку ряд отобранных факторов состояли из большого количества пунктов опросника, специалисты INPO решили дополнительно применить метод главных компонент для этих факторов с целью выделения в них субфакторов. В итоге было получено 14 субфакторов. Какое-либо описание этой процедуры отсутствует.

В отчете также указано, что многие респонденты не ответили на все утверждения анкеты (пропущенные данные) или отвечали выбором варианта “не знаю”, что может указывать на то, что респонденты были сбиты с толку утверждением или не находили его соответствующим для своей работы.

Необходимо учитывать, что модели анализа главных компонент имеют ряд ограничений. Одно из них состоит в том, что совершенно не очевидно, как поступать в ситуации, когда мы имеем дело с неполным набором данных или с пропущенными данными (Chen H., 2002). На сегодняшний день, неполные данные либо исключаются, либо заполняются с использованием различных методов интерполяции. При этом, если переменные не отвечают требованию нормального распределения, мы не можем использовать средние оценки для интерполяции пропущенных данных. Однако методы интерполяции не действуют, когда значимая часть матрицы измерений неизвестна. Из данных, представленных INPO, неясно, какой метод использовался для заполнения пропущенных данных при выполнении PCA и какая часть матрицы измерений была неизвестна.

После факторного анализа методом главных компонент ответов на полный список утверждений опросника ($N = 110$), специалисты INPO решили сократить число пунктов опросника до 60 за счет пропущенных данных и неопределенных ответов, а также пунктов, имеющих слабую корреляцию со всеми другими. При этом число выделенных общих факторов сократилось с 9 до 7 (анализ субфакторов не упоминается). К сожалению, в этом случае специалисты INPO не только не обосновали полученное факторное решение, но даже не привели общий процент дисперсии данных, который приходится на эти 7 факторов.

Таким образом, в представленном виде невозможно использовать результаты анализа данных опроса методом главных компонент (PCA) для выделения и обоснования 9 общих факторов и 14 субфакторов размерности конструкта культуры безопасности.

Метод главных факторов

Другой способ оценки факторной структуры, распространенный в области социальных наук, является метод главных факторов или метод главных осей (Principal Axis Factoring - PAF). Основное различие двух моделей факторного анализа состоит в том, что в анализе главных компонент предполагается, что должна быть использована вся изменчивость переменных, тогда как в анализе главных факторов используется только изменчивость переменной, общая и для других переменных (StatSoft, 1999). В большинстве случаев эти два метода приводят к весьма близким результатам. Однако анализ главных компонент часто более предпочтителен как метод сокращения данных, в то время как анализ главных факторов лучше применять с целью определения структуры данных (StatSoft, 1999). Метод главных факторов обладает более мощным и более сложным аппаратом, и является более чувствительным методом для выявления латентной структуры исследуемого конструкта (Ford et al., 1986), как культура безопасности в нашем случае.

С этих позиций специалисты NRC решили выполнить независимый анализ главных факторов по данным опроса персонала АЭС, предоставленным INPO. Они выполнили серию PAF для ответов респондентов на все 110 пунктов опросника и различные другие комбинации пунктов: с удалением групп элементов с низкими корреляциями между пунктами, с удалением анкет с чрезмерно недостающими данными, с удалением пунктов с низкими факторными нагрузками и пунктов, которые имели высокую корреляцию с более чем одним фактором. Во всех случаях анализ главных факторов дал от 7 до 9 интерпретируемых факторов.

К сожалению, в отчете отсутствует обоснование условий для выполнения метода главных факторов (Лоули Д.Н., Максвелл А.Э., 1967; Иберла К., 1980; Ким Дж.-О. и др., 1989; Rencher A.C., 2002; Машин В.А., 2010):

- Проверка допущения о многомерном нормальном распределении исходных переменных;
- Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных;
- Выбор метода вращения главных осей;
- Проверка целесообразности применения факторной модели анализа к исходным данным;
- Выполнение процедуры отбора общих факторов;
- Оценка значимости полученного факторного решения с помощью статистических критериев и принципа "простой структуры".

Как и в случае анализа главных компонент, из представленного отчета неясно, какой тип корреляции был использован при анализе главных факторов. Поскольку это специально не проговаривается, можно вновь предположить, что речь идет о стандартном использовании коэффициента корреляции Пирсона. Как и ранее приходится заметить, что применение корреляции Пирсона для порядковых данных является не корректным, и для матрицы парных корреляций следует использовать ранговые коэффициенты корреляции Спирмена или Кендалла. Также в отчете отсутствуют матрица факторных нагрузок, матрица собственных значений и матрица значений общности, не указан суммарный процент дисперсии выделенных факторов; все это является необходимым для понимания и обоснования полученного факторного решения. Кроме этого, неясно, какой метод использовался для заполнения пропущенных данных при выполнении PAF.

Таким образом, в представленном виде невозможно использовать результаты анализа главных факторов (PAF) для обоснования полученных факторных решений и делать выводы о размерности конструкта культуры безопасности на основании ответов респондентов.

Коэффициент альфа Кронбаха

Для вычисления надежности опросника по оценке однородности или внутренней согласованности пунктов опросника культуры безопасности специалисты NRC использовали коэффициент альфа Кронбаха (который приложим к порядковым данным, полученным на основе шкалы Лайкерта). Данная статистика рассчитывалась для факторного решения, полученного INPO для полного варианта опросника из 110 пунктов: 9 факторов и 14 субфакторов. При этом использовались данные респондентов по сокращенному варианту опросника из 60 пунктов. Обоснование такого решения в отчете отсутствует.

При расчете коэффициента альфа Кронбаха необходимо учитывать, что если множество анализируемых пунктов опросника имеет несвязанные между собой подмножества пунктов, с высокой внутренней корреляцией (т.е. подмножества пунктов измеряют разные характеристики или факторы), значение коэффициента альфа Кронбаха будет высоким (Stahl D., 2007). Вот почему коэффициент альфа Кронбаха должен применяться только после того, как выполнен факторный анализ и получено обоснованное факторное решение, отражающее латентную структуры исследуемого конструкта (в нашем случае – культуры безопасности). Но результаты анализа главных компонент и главных факторов, которые представлены в отчете, не позволяют использовать их для определения структуры исследуемого конструкта культуры безопасности, следовательно, отсутствует условие использования коэффициента альфа Кронбаха для оценки внутренней согласованности пунктов опросника по каждому фактору и субфактору.

Внутриклассовый коэффициент корреляции

Для определения внутригрупповой надежности (степени близости ответов на пункты опросника культуры безопасности по респондентам одной АЭС) в отчете использовался коэффициент внутриклассовой (внутригрупповой) корреляции (ICC - Intraclass Correlation Coefficient) (McGraw K.O., Wang S.P., 1996; Doros G., Lew R., 2010). Отличие его от коэффициента Пирсона состоит в том, что при расчете ICC проверяется также связанность средних уровней переменных.

Важно помнить, что при использовании статистики ICC должны быть рассмотрены следующие основные допущения (Stahl D., 2007):

- Равенство значений дисперсий переменных;
- Многомерное нормальное распределение данных;
- Интервальный характер данных;
- Независимость выборок (т.е., результаты опроса одного человека должны быть независимы от результатов опроса другого человека).

Из отчета неясно, была ли выполнена проверка равенства дисперсий (на вопросах нормальности распределения мы остановимся ниже), но тот факт, что данные опроса имели порядковый характер (шкала Лайкерта), а не интервальный делает невозможным применение коэффициента внутриклассовой корреляции (ICC) для оценки внутригрупповой надежности результатов исследования.

Описательные статистики

Специалисты NRC выполнили описательный анализ факторов культуры безопасности на основе данных исследования INPO, результаты которого представили в таблице. В нее вошли данные по средним арифметическим и стандартным отклонениям ответов респондентов по каждому из 9 факторов и опроснику культуры безопасности в целом, а также минимумы и максимумы.

Поскольку данные ответов респондентов получены с помощью порядковой шкалы Лайкерта, в качестве средних можно использовать медиану, но не среднее арифметическое. (Орлов А.И., 2004). Точно также для описания порядковой переменной в качестве меры рассеивания (отклонения от среднего) мы не можем использовать стандартное отклонение или дисперсию, а должны воспользоваться вариационным размахом или квантильным рангом (Мангейм Дж.Б., Рич Р.К., 1997).

Взаимная корреляция факторов

Специалисты NRC рассчитали парные коэффициенты корреляции Пирсона для 9 факторов культуры безопасности, предполагая, что между ними должна быть высокая

взаимная корреляция, поскольку они представляют разные грани одного исходного конструкта (культуры безопасности). В то же время они не должны быть настолько тесно связанными, что противоречило бы многомерности конструкта культуры безопасности, и указывало бы на то, что факторы измеряют, по сути, одно и то же. Значения парных корреляций в диапазоне $0.70 \div 0.95$, по мысли специалистов NRC, должно было обеспечить одновременно взаимосвязь и уникальность факторов культуры безопасности. Обоснование данного предположения отсутствует.

Необходимо напомнить, что метод главных компонент направлен на выделение из заданного множества переменных (в нашем случае это ответы респондентов на пункты опросника культуры безопасности) подмножеств переменных, тесно связанных между собой, которые образуют явно не наблюдаемые, латентные факторы, отражающие глубинную структуру изучаемого конструкта (Митина О.В., Михайловская И.Б., 2001). При этом переменные, входящие в одно подмножество, в значительной степени независимы от переменных из других подмножеств.

Другими словами, главные компоненты представляют собой линейную комбинацию наблюдаемых переменных, обладающую свойством ортогональности (Ким Дж.-О. и др., 1989). Ортогональность (некоррелируемость) факторов не означает, что между факторами не существует никакой связи. Достаточно, чтобы связь между ними была нелинейной (Круг Г.К., 1983). Нарушение свойства ортогональности или сильная корреляция между факторами выражается в мультиколлинеарности (линейная связь между более чем двумя объясняющими переменными), которая ведет к неустойчивости оценок параметров факторной модели и сложности интерпретации влияния тех или иных факторов на изучаемый конструкт. Основной причиной возникновения мультиколлинеарности может служить то, что выделенные в ходе анализа главных компонент факторы характеризуют одну и ту же сторону изучаемого конструкта.

Одним из индикаторов присутствия мультиколлинеарности между факторами является превышение значений парных коэффициентов корреляции по абсолютной величине $0.75 \div 0.80$. Авторы отчета представили таблицу парных коэффициентов корреляции между факторами культуры безопасности ($F_1 \div F_9$, см. таблицу 1), высокие значения которых позволяет предположить присутствие мультиколлинеарности. Чтобы убедиться в этом, мы рассчитали значение определителя (детерминанта) корреляционной матрицы. Близость значений определителя матрицы к нулю свидетельствует о наличии мультиколлинеарности (близость к 1 – об ее отсутствии) (Айвазян С.А., Мхитарян В.С., 1998). Для матрицы парных корреляций 9 факторов культуры безопасности значение определителя составило 3.05×10^{-5} , что с высокой вероятностью указывает на наличие мультиколлинеарности и возвращает нас к

вопросу относительно обоснованности факторного анализа и критериев отбора главных компонент.

	F ₂	F ₃	F ₄	F ₅	F ₆	F ₇	F ₈	F ₉
F ₁	0.81							
F ₂	0.93	0.85						
F ₃	0.81	0.71	0.81					
F ₄	0.83	0.76	0.80	0.69				
F ₅	0.89	0.83	0.89	0.82	0.85			
F ₆	0.26	0.45	0.37	0.28	0.37	0.46		
F ₇	0.86	0.79	0.89	0.75	0.77	0.85	0.28	
F ₈	0.76	0.64	0.75	0.66	0.70	0.80	0.48	0.72

Таблица 1. Взаимные корреляции факторов культуры безопасности.

Критериальный тип валидности

Статистикой, которой специалисты NRC воспользовались для проверки взаимосвязи между факторами культуры безопасности и показателями эффективности деятельности АЭС (критериальный тип валидности), был коэффициент корреляции Пирсона. Известно, что коэффициент корреляции Пирсона является параметрическим критерием, который проверяет линейную зависимость между двумя переменными. Корреляция Пирсона предназначена для использования, когда обе переменные измерены либо интервальной шкалой, либо шкалой отношений, и при этом каждая переменная нормально распределена. Переменные культуры безопасности (ответы респондентов на утверждения опросника) измерены порядковой шкалой Лайкерта. Для порядковых данных, согласно общих требований (Сидоренко Е.В., 2003; Бащинский С.Е., Власов В.В., 2006; Ланг Т.А., Сесик М., 2011), необходимо использовать ранговые коэффициенты корреляции Спирмена или Кендалла.

Использование параметрического коэффициента корреляции Пирсона для порядковых переменных в качестве оценки критериальной валидности опросника культуры безопасности является некорректным.

Проверка нормальности распределений переменных

Специалисты NRC выполнили проверку предположения о нормальности распределения для 9 факторов культуры безопасности и показателей эффективности деятельности АЭС с помощью критерия Шапиро-Уилка. Ими было установлено, что факторы культуры безопасности распределены нормально, а показатели эффективности деятельности АЭС нет. При этом в отчете не приведены значения критерия Шапиро-Уилка (W) для переменных и соответствующие уровни значимости (p). Читателю приходится на слово

поверить авторам при условии, что ненормальные распределения скорее являются нормой в реальном мире (Pyzdek T., 1999).

Заметим, что использование непараметрических коэффициентов корреляции Спирмена или Кендалла, которые не требуют выполнения предположения о нормальности распределения переменных, позволило бы корректно оценить критериальную валидность опросника культуры безопасности относительно показателей эффективности деятельности АЭС. В случае наличия одинаковых рангов в переменной (например, несколько АЭС имеют одинаковое количество незапланированных срабатываний аварийной защиты, на что указывают авторы отчета) рекомендуется использовать коэффициент корреляции Спирмена с поправкой на одинаковые ранги при расчете (Суходольский Г.В., 1998).

Анализ коэффициентов корреляции

Специалисты NRC приводят значения коэффициентов корреляции Пирсона (r) между 9 факторами культуры безопасности и показателями эффективности деятельности АЭС, которые оказались в диапазоне от -0.25 до -0.48. При этом коэффициенты корреляции, согласно авторам отчета, в диапазоне от -0.25 до -0.30 должны отражать слабые и умеренные влияния факторов культуры безопасности на показатели эффективности деятельности АЭС, а более высокие, по абсолютной величине, значения корреляции могут указывать на явные влияния. Дополнительно в отчете представлены результаты мета-анализа исследований влияния культуры безопасности на количество происшествий; значения коэффициентов корреляции (r) между этими переменными оказались в диапазоне от -0.22 до -0.39.

Следует напомнить, что статистическая значимость (p) коэффициента корреляции зависит от объема выборки (Титкова Л.С., 2002). Чем больше объем выборки, тем выше статистическая значимость при одном и том же коэффициенте корреляции. В результате, при больших объемах выборки слабая корреляция может оказаться статистически значимой (примеры из отчета: $r = 0.048$, $p = 0.01$, $N = 2876$; $r = 0.037$, $p = 0.05$, $N = 2876$). Для разрешения этой коллизии рекомендуют, в первую очередь, проанализировать диаграмму рассеяния двух переменных. Во-вторых, воспользоваться рекомендациями, предлагаемыми в руководствах к статистическому пакету для социальных наук SPSS, одному из лидеров в области статистической обработки информации (Бююль А., Цефель П., 2005; Дорогонько Е.В., 2010):

$0.0 < r \leq 0.2$	Очень слабая корреляция
$0.2 < r \leq 0.5$	Слабая корреляция
$0.5 < r \leq 0.7$	Средняя корреляция
$0.7 < r \leq 0.9$	Сильная корреляция
$0.9 < r \leq 1.0$	Очень сильная корреляция

Как можно видеть, согласно рекомендациям руководств SPSS, все представленные в отчете коэффициенты корреляции Пирсона (при том, что их использование было некорректным для порядковых и нормально нераспределенных данных) указывают на слабую связь 9 факторов культуры безопасности с показателями эффективности деятельности.

Также важно понимать, что коэффициент корреляции (r) между двумя переменными не указывает на причинно-следственную связь между ними. Когда две переменные высоко коррелируют, это может быть связано с наличием у них некоторого общего причинного фактора, поэтому нельзя автоматически полагать, что изменения одной переменной влияют на изменения другой переменной (Аткинсон Р.Л. и др., 1999). Кроме этого, при наличии корреляции не всегда очевидно, что является причиной, а что следствием. Например, высокая корреляция между чувством собственного достоинства и академическими успехами предполагает два варианта причинно-следственной связи: либо академические успехи ведут к высокому самоуважению, либо чувство собственного достоинства влияет на успехи в учебе (Майерс Д., 1997).

Если мы воспользуемся моделью парной линейно регрессии для анализа причинно-следственных связей при взаимодействии факторов культуры безопасности с показателями эффективности деятельности ($Y_x = a + bx$, где Y - показатель эффективности деятельности АЭС, а x - фактор культуры безопасности), то в этом случае, квадрат коэффициента корреляции Пирсона r_{yx}^2 (называемый коэффициентом детерминации или R^2) будет определять долю дисперсии переменной Y , вызванную влиянием переменной x (Елисеева И.И. и др., 2003). Представленные в отчете значения коэффициентов корреляции в диапазоне от -0.25 до -0.48 означают, что от 6.3% ($R^2 = 0.063$) до 20.3% ($R^2 = 0.203$) дисперсии показателей эффективности деятельности АЭС можно объяснить влиянием факторов культуры безопасности. При этом на долю прочих факторов приходится от 93.7% до 79.7% дисперсии показателей эффективности деятельности АЭС. Исходя из модели парной линейной регрессии, о влиянии зависимой переменной (x) на независимую (Y) можно говорить лишь в том случае, когда коэффициент детерминации (R^2) равен не менее 0.5, а значит $r_{yx} > 0.7$ (Бююль А., Цефель П., 2005; Дорогонько Е.В., 2010). Величина коэффициента детерминации служит одним из критериев оценки качества линейной модели регрессии. Чем больше доля объясненной вариации, тем меньше роль прочих факторов и, следовательно, линейную модель регрессии можно использовать для прогноза значений независимой переменной (Елисеева И.И. и др., 2003). Расчет коэффициентов детерминации для корреляций, представленных в отчете (даже не принимая во внимание статистическую некорректность при определении факторов), не позволяет использовать линейные модели регрессии для прогноза показателей эффективности деятельности АЭС по значениям факторов культуры безопасности.

Диаграммы рассеяния

В заключение авторы отчета привели несколько диаграмм рассеяния общих оценок культуры безопасности ("*Safety Culture Overall*") и отдельных показателей эффективности деятельности АЭС ($N=63$). Одна из диаграмм представлена на рис. 1, где "*SCWE-Related Allegations in 2010*" - показатель количества жалоб персонала АЭС, которые он направил в адрес NRC в 2010 году по программе "Создания Рабочей Обстановки с Сознательным Отношением к Вопросам Безопасности" (Safety Conscious Work Environment) (NRC, 2005). На графике также построена линия линейной регрессии и приведено значение коэффициента детерминации (R^2). Такие графики очень удобны для визуальной проверки выбросов по диаграмме рассеяния и линейного характера связи между переменными.

Выбросами считаются атипичные или аномальные наблюдения, не согласующиеся с остальными элементами выборки наблюдений. Появление выбросов связано с влиянием на переменную не учитываемых факторов, а также ошибками на стадии измерения переменной. Коэффициент корреляции Пирсона неустойчив к выбросам (в отличие от коэффициентов корреляции Спирмена и Кендалла), поэтому при его расчетах все аномальные наблюдения должны быть исключены (Шишлянникова Л.М., 2009). Это относится и к факторному анализу, который в качестве исходных данных использует матрицу парных коэффициентов корреляции Пирсона; наличие выбросов негативно влияет на точность конечного факторного решения. Атипичные наблюдения резко влияют на результаты регрессионного анализа и могут приводить к искажениям при построении линии регрессии и грубым ошибкам предсказания значений исследуемой переменной по оценкам зависимой переменной (например, показателей эффективности деятельности по уровню культуры безопасности).

Для статистической оценки выбросов разработаны многочисленные методы (Кобзарь А.И., 2006). Но первый шаг, это визуальный анализ распределения переменных. Наблюдение, которое на рис.1 обведено красным кружком, имеет явно аномальный характер и вносит существенное искажение как в значение коэффициента корреляции Пирсона (если $R^2 = 0.078$, то $r = 0.28$), так и в модель линейной регрессии (согласно которой, лишь 7.8% дисперсии общих оценок культуры безопасности объясняется влиянием количества жалоб, которые персонал направил в адрес NRC в 2010 году). Исключение данного атипичного наблюдения с очевидностью приведет к значительному снижению как значения коэффициента корреляции (r), так и коэффициента детерминации (R^2).

Представленные в отчете диаграммы рассеяния указывают на то, что в ходе корреляционного и факторного анализов не был выполнен анализ на атипичные наблюдения, который подразумевает исключение выбросов, а также исследование возможных причин отклонений от выборочных данных.

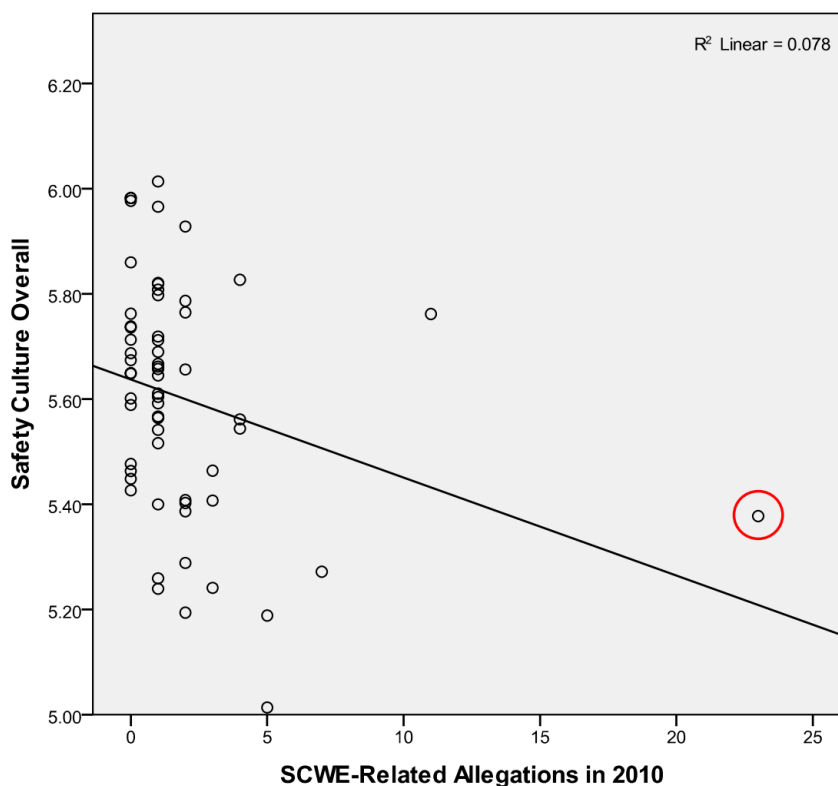


Рисунок 1. Диаграмма рассеяния общих оценок культуры безопасности и показателя эффективности деятельности АЭС.

Заключение

Сформулируем основные выводы, к которым мы пришли при рассмотрении независимой оценки NRC результатов исследования культуры безопасности INPO.

Выбор INPO минимального объема выборки в 30 человек с каждой АЭС для проведения анкетирования по опроснику культуры безопасности не отвечает требованиям репрезентативности и не позволяет выборочным наблюдениям с приемлемой точностью отражать характеристики генеральной совокупности.

Является некорректным использование порядковой шкалы Лайкерта в качестве интервальной, с последующим применением к выборочным данным статистических критериев, разработанных для интервальных шкал.

При проведении факторного анализа методом главных компонент (РСА) отсутствовало описание метода для заполнения пропущенных данных; проверка допущения о многомерном нормальном распределении исходных переменных, выбор типа матрицы

парных коэффициентов корреляции в качестве исходных данных; проверка целесообразности применения факторной модели анализа к исходным данным; выполнение процедуры отбора общих факторов; матрица факторных нагрузок, матрица собственных значений и матрица значений общности, что является необходимым для понимания и обоснования полученного факторного решения.

При использовании метода главных факторов (PAF) отсутствовало описание метода для заполнения пропущенных данных; проверка допущения о многомерном нормальном распределении исходных переменных, выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных; выбор метода вращения главных осей; проверка целесообразности применения факторной модели анализа к исходным данным; выполнение процедуры отбора общих факторов; оценка значимости полученного факторного решения с помощью статистических критериев и принципа "простой структуры"; суммарный процент дисперсии выделенных факторов; матрица факторных нагрузок, матрица собственных значений и матрица значений общности, что является необходимым для понимания и обоснования полученного факторного решения.

Поскольку результаты анализа главных компонент и главных факторов, которые представлены в отчете, не позволяют использовать их для определения структуры исследуемого конструкта культуры безопасности, нельзя воспользоваться коэффициентом альфа Кронбаха для оценки внутренней согласованности пунктов опросника по каждому фактору культуры безопасности.

При использовании коэффициента внутриклассовой корреляции (ICC) отсутствовала проверка равенства значений дисперсий переменных и многомерного нормального распределения данных. Было нарушено ограничение на использование ICC для интервальных данных.

Описательная статистика факторов культуры безопасности содержит значения средних и стандартных отклонений. Поскольку данные ответов респондентов получены с помощью порядковой шкалы Лайкерта, в качестве средних можно использовать медиану, а в качестве меры рассеивания - вариационный размах или квантильный ранг.

При рассмотрении взаимной корреляции факторов отсутствует анализ мультиколлинеарности, которая ведет к неустойчивости оценок параметров факторной модели и сложности интерпретации влияния тех или иных факторов на изучаемый конструкт. Выполненная оценка мультиколлинеарности с помощью расчета определителя матрицы парных корреляций указывает на высокую вероятность наличия выраженных линейных связей между факторами и требует анализа обоснованности факторного решения и критериев отбора главных компонент.

Использование параметрического коэффициента корреляции Пирсона в качестве оценки критериальной валидности опросника культуры безопасности, когда одна переменная носит порядковый характер, а другая имеет выраженные отклонения от нормального распределения, является некорректным.

В отчете не приведены значения критерия Шапиро-Уилка (W) и уровни значимости (p) при проверке предположения о нормальности распределения факторов культуры безопасности и показателей эффективности деятельности.

В ходе корреляционного и факторного анализов не был выполнен анализ на атипичные наблюдения, который подразумевает исключение выбросов, а также исследование возможных причин отклонений от выборочных данных. Наличие аномальных наблюдений могло значительно завязать значения коэффициентов корреляции Пирсона, а также исказить результаты факторного анализа.

Некорректное использование методов статистического анализа результатов исследования культуры безопасности персонала АЭС США не позволяет сделать статистически обоснованные выводы относительно выполненной оценки NRC валидности и надежности опросника культуры безопасности, разработанного INPO. Вероятно, исследователи столкнулись с проблемой обработки больших объемов данных и ограниченными ресурсами (человеческими, финансовыми) для выполнения этих работ. Но данные моменты не могут служить оправданием слабой культуры статистического анализа и не соблюдением требований в области математической статистики.

Список литературы

1. Леонов В.П., Ижевский П.В. Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям // Бюллетень ВАК РФ, 1997. Вып. 5, с.56-61.
2. Гланц С. Медико-биологическая статистика. М., Практика, 1998.
3. Леонов В.П. Применение статистики в статьях и диссертациях по медицине и биологии. Часть IV. Наукометрия статистической парадигмы экспериментальной медицины // Международный журнал медицинской практики, 2002а. № 3, с. 6-11.
4. Ланг Т.А. Двадцать ошибок статистического анализа, которые вы сами можете обнаружить в биомедицинских статьях // Международный журнал медицинской практики, 2005. № 1, с. 21-31.
5. Леонов В.П. Наукометрия статистической парадигмы экспериментальной биомедицины (по материалам публикаций) // Вестник Томского государственного университета. Серия "Математика. Кибернетика. Информатика", 2002б. № 275, с. 17-24.
6. Сидоренко Е.В. Методы математической обработки в психологии. С-Пб.: Речь, 2003.

7. Бащинский С.Е., Власов В.В. (ред.). Рекомендации по подготовке научных медицинских публикаций. Сборник статей и документов. М.: Медиа Сфера, 2006.
8. Ланг Т.А., Сесик М. Как описывать статистику в медицине. Руководство для авторов, редакторов и рецензентов. М.: Практическая медицина, 2011.
9. Гласс Дж., Стэнли Дж. Статистические методы в психологии и педагогике. М.: Прогресс, 1976.
10. Бащинский С.Е. Статистика умеет много гитик // Международный журнал медицинской практики, 1998. № 4, с. 13-15.
11. IAEA. Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident. Safety Series, No 75-INSAG-1. IAEA, 1986.
12. IAEA. Safety Culture. Safety Series, No 75-INSAG-4. IAEA, 1991.
13. IAEA. Nuclear Power Plant Personnel Training and its Evaluation. A Guidebook. Technical Reports Series, No 380. IAEA, 1996.
14. МАГАТЭ. Руководство ASCOT: Руководство по самостоятельной оценке культуры безопасности и проведению миссии ASCOT. TECDOC-743/R. МАГАТЭ, 1994.
15. МАГАТЭ. Самооценка культуры безопасности на предприятиях атомной промышленности. Основные положения и положительная практика. TECDOC-1321. МАГАТЭ, 2002.
16. ВАО АЭС. Повреждение крышки реактора на АЭС Дэйвис-Бесси. Сообщение о важном опыте эксплуатации. WANO SOER 2003-2. 2003.
17. INPO. Principles for a Strong Nuclear Safety Culture. INPO, 2004.
18. WANO. Principles for a Strong Nuclear Safety Culture. WANO Guideline. GL 2006-02. WANO, 2006.
19. IAEA. Application of the management system for facilities and activities. Safety guide. Safety Standards Series, No GS-G-3.1. IAEA, 2006.
20. IAEA. SCART Guidelines: Reference report for IAEA Safety Culture Assessment Review Team (SCART). Services Series, No 16. IAEA, 2008.
21. NRC. Final Safety Culture Policy Statement. Nuclear Regulatory Commission. Federal Register, Vol. 76, No. 114. 2011.
22. SCK/CEN. Safety Culture Assessment Tools in Nuclear and Non-Nuclear Domains. Review of safety culture tools. Open report of the Belgian Nuclear Research Centre. SCK/CEN, 2012.
23. Keil Centre. Managing Safety Culture in the UK Rail Industry. Report on the Review of Safety Culture Tools and Methods. The Keil Centre Ltd., 2003.
24. HSE. A review of safety culture and safety climate literature for the development of the safety culture inspection toolkit. Research Report 367. Health and Safety Executive, 2005.

25. Singla A.K. Assessing Patient Safety Culture: A Review and Synthesis of the Measurement Tools. *Journal of Patient Safety*, 2006. Vol. 2, No 3. p. 105-115.
26. EUNETPAS. Use of Patient Safety Culture Instruments and Recommendations. The European Network for Patient Safety, 2010.
27. Health Foundation. Evidence scan: Measuring safety culture. The Health Foundation, UK. 2011.
28. Fricke-Ernst C., Kluge A. Overview on safety culture & safety climate assessment – a closer look on the culture, the climate and the safety aspect. Proceedings of the SSCAOC workshop, Antwerp, Belgium. 2012.
29. SNL. A Literature Review of Safety Culture. Sandia National Laboratories, DOE. 2013.
30. NLR. Safety Methods Database. National Aerospace Laboratory of the Netherlands, 2013.
31. Morrow S., Barnes V. Independent Evaluation of INPO's Nuclear Safety Culture Survey and Construct Validation Study. (<http://pbadupws.nrc.gov/docs/ML1217/ML12172A093.pdf>)
32. Ивановский Р.И. Теория вероятностей и математическая статистика. Основы, прикладные аспекты с примерами и задачами в среде Mathcad. СПб.: БХВ-Петербург, 2008.
33. Сенатов В.В. Центральная предельная теорема: Точность аппроксимации и асимптотические разложения. М.: Эдиториал УРСС, 2009.
34. Москвин С. Выборка в социологическом исследовании // Военно-социологическое исследование, 1993. с. 27-30.
35. Гончаров А.И. Методы изучения общественного мнения и социальной информации. М.: ГА ВС, 1993.
36. Носс И.Н. Введение в технологию психодиагностики. М.: Изд-во Института Психотерапии, 2003.
37. Bland M. An Introduction to Medical Statistics. Third Edition. Oxford University Press, Oxford. 2000.
38. Lenth R.V. Some Practical Guidelines for Effective Sample Size Determination // *The American Statistician*, 2001. Vol. 55, No. 3. p. 187-193.
39. Бурлачук Л.Ф., Морозов С.М. Словарь-справочник по психодиагностике. СПб.: Питер-Ком, 1999.
40. Малхорта Н.К. Маркетинговые исследования. Практическое руководство. М.: Издательский дом "Вильямс", 2002.
41. Мещеряков Б.Г., Зинченко В.П. (сост., общ. ред.) Большой психологический словарь. М.: ОЛМА-ПРЕСС, 2003.
42. Bond T.G., Fox C.M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Psychology Press, 2013
43. Лоули Д.Н., Максвелл А.Э. Факторный анализ как статистический метод. М.: Мир. 1967.
44. Иберла К. Факторный анализ. М.: Статистика. 1980.

45. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р., Олдендерфер М.С., Блэшфилд Р.К. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика. 1989.
46. Rencher A.C. Methods of multivariate analysis. Wiley. 2002.
47. Chen H. Principal component analysis with missing data and outliers. Technical report. Rutgers University, 2002.
48. StatSoft. Фирменное руководство по STATISTICA 6.0. Том III. Факторный анализ. StatSoft, 1999.
49. Ford J., MacCallum R., Tait M. The application of exploratory factor analysis in applied psychology: A critical review and analysis // Personnel Psychology, 1986. No 39, p. 291-314.
50. Машин В.А. Методические вопросы использования факторного анализа на примере спектральных показателей сердечного ритма // Экспериментальная психология, 2010. № 4. с. 119-138.
51. Stahl D. Introduction to measurement and scale development. Department of Biostatistics & Computing. King's College London. 2007.
52. McGraw K.O., Wang S.P. Forming inferences about some intraclass correlation coefficients // Psychological Methods, 1996. Vol. 1(1), p. 30-46.
53. Doros G., Lew R. Design Based on Intra-Class Correlation Coefficients // American Journal of Biostatistics, 2010. Vol. 1 (1), p. 1-8.
54. Орлов А.И. Нечисловая статистика. М.: МЗ-Пресс, 2004.
55. Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. М.: Изд-во «Весь Мир», 1997.
56. Митина О.В., Михайловская И.Б. Факторный анализ для психологов. М.: УМК «Психология», 2001.
57. Круг Г.К. (ред.) Статистические методы в инженерных исследованиях. Учебное пособие. М.: Высшая школа, 1983.
58. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.
59. Pyzdek T. Non-Normal Distributions in the Real World. Quality Digest, 1999. p. 36-41.
60. Суходольский Г.В. Основы математической статистики для психологов. СПб.: Изд-во Санкт-Петербургский университет, 1998.
61. Титкова Л.С. Математические методы в психологии. Владивосток: Изд-во Дальневосточного университета, 2002.
62. Бююль А., Цефель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. ДиаСофт, 2005.
63. Дорогонько Е.В. Обработка и анализ социологических данных с помощью пакета SPSS. Сургут: Издательский центр СурГУ, 2010.

64. Аткинсон Р.Л., Аткинсон Р.С., Смит Э.Е., Бем Д.Д., Нолен-Хоэксема С. Введение в психологию. М.: Тривола, 1999.
65. Майерс Д. Социальная психология. СПб.: Питер, 1997.
66. Елисеева И.И., Курышева С.В., Костеева Т.В. и др. Эконометрика. М.: Финансы и статистика, 2003.
67. NRC. Guidance for Establishing and Maintaining a Safety Conscious Work Environment. NRC Regulatory Issue Summary 2005-18. NRC, 2005.
68. Шишлянникова Л.М. Применение корреляционного анализа в психологии // Психологическая наука и образование, 2009. № 1, с. 98 - 107.
69. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006.