

Машин В.А.

Комментарий к технической статье

Независимая оценка результатов анкетирования INPO по вопросам культуры безопасности и анализ конструктивной валидности. 2012

Independent Evaluation of INPO's Nuclear Safety Culture Survey and Construct Validation Study. Stephanie Morrow, Valerie Barnes

<http://pbadupws.nrc.gov/docs/ML1217/ML12172A093.pdf>

Машин В.А.

Главный специалист Центрального Института Повышения Квалификации (ЦИПК, Обнинск); канд. психологических наук. E-mail: mashin-va@mail.ru

Краткое пояснение

Опросник для оценки культуры безопасности персонала АЭС был разработан INPO¹ и в первоначальном виде содержал 110 пунктов (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 0 выбор для "не знаю, не имею возможности наблюдать".

В опросе участвовал персонал 63 станций США. К участию было приглашено 6333 человек, в итоге было получено 3031 анкеты, из которых было отобрано 2876, которые были полностью заполнены по 110 пунктам, или почти полностью. В среднем, от каждой станции, кроме двух, было получено не менее 30 анкет. После первичного анализа методом главных компонент (РСА), было принято решение использовать в дальнейшем анализе лишь результаты по 60 пунктам опросника.

Опросник был разработан по заказу Института ядерной энергетики США (Nuclear Energy Institute - NEI) для оценки культуры безопасности персонала АЭС США. Основными целями исследования было изучение факторов, которые определяют культуру безопасности в области ядерной энергетики, и оценка взаимосвязи между факторами культуры безопасности, выявленными с помощью опросника, и другими показателями безопасности и эффективности организации.

Замечание

Фрагменты статьи выделены голубым фоном.

¹ Institute of Nuclear Power Operations - Институт по эксплуатации АЭС (США)

2.4. Взаимосвязь между культурой безопасности и показателями безопасности

Результаты мета-анализа исследований в этой области обнаружили непротиворечащие друг другу доказательства статистической значимости линейной зависимости между культурой безопасности и количеством аварий/травм, со значением коэффициента корреляции в диапазоне от **-.22** до **-.39** ($p < .05$).

Согласно Джейкобу Коэну (J. Cohen), взаимосвязь между культурой безопасности и показателями безопасности попадает в область **средних** значений.

Согласно Achim Bühl, Peter Zöfel (2001):

$0.0 < r \leq 0.2$	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

См. дополнительно:

- Achim Bühl; Peter Zöfel SPSS Version 10: Einführung in die moderne Datenanalyse unter Windows. 2001.

3.2. Организация анкетирования

(Определение размера выборки)

Центральная Предельная Теорема утверждает, что по мере того, как размер случайной выборки возрастает, выборочное среднее приближается к среднему генеральной совокупности (Myers & Hy, 2003). Когда размер случайной выборки становится больше **30**, она, обычно, приблизительно имеет нормальное распределение и ее среднее значение незначительно отличается от среднего значения генеральной совокупности. Таким образом, размер выборки, равный **30**, обычно используется в качестве минимального порога, который устанавливается, чтобы выборка представляла собой хорошую оценку для генеральной совокупности.

Центральная Предельная Теорема не используется для определения размера выборки в исследованиях генеральной совокупности. 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.

- ✓ Мы должны учитывать при определении размеры выборки:
- Доверительный интервал (задает ширину диапазона по обе стороны от выбранной точки, куда могут попадать ответы)
 - Доверительная вероятность (показывает, с какой вероятностью ответ попадет в границы доверительного интервала)
 - Размер генеральной совокупности

См. дополнительно:

Калькулятор размера выборки (Creative Research Systems)

<http://www.surveysystem.com/sscalc.htm>

Пример:

- Доверительная вероятность = 0.95
- Доверительный интервал = 5
- Генеральная совокупность = 600
- Необходимый размер выборки = **234**

➤ **Размер выборки не удовлетворяет статистическим требованиям**

См. также:

- Bland M. (2000) An Introduction to Medical Statistics (*Sample Size*)
- Russell V. LENTH Some Practical Guidelines for Effective Sample Size Determination. The American Statistician, Vol. 55, No. 3. (Aug., 2001)

3.1. Разработка пунктов опросника для проведения анкетирования

(Тип данных)

Участников обследования просили ответить на пункты опросника, оценив степень согласия с каждым утверждением по 7-балльной шкале Лайкерта, проранжированной от выраженного несогласия до выраженного согласия. Система баллов от 1 до 7 позволяет исследователю приписать количественные значения для качественных оценок респондентов. Эти данные традиционно

используются в качестве приближения к **интервальной шкале**. Это означает, что варианты ответов располагаются на относительно равном расстоянии друг от друга по шкале оценок.

Шкала Лайкерта является **порядковой шкалой** (или ранговой), 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. 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. Особенностью шкалы интервалов является то, что нулевая точка здесь выбирается произвольно:

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

➤ Некорректно использование шкалы Лайкерта в качестве интервальной.

См. примеры в работе:

- Bond, T. G., & Fox, C. M. (2001) Applying the Rasch Model: Fundamental Measurement in the Human Sciences.

4.1. Исследовательский факторный анализ

(Анализ Главных Компонент - Principal Components Analysis, PCA)

Подход INPO к определению размерности культуры безопасности, на основании оценок для утверждений, включенных в анкетирование по вопросам культуры безопасности, базировался на **анализе главных компонент (PCA)**.

✓ Для использования PCA не было рассмотрено:

- допущение 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.

See more

- Rencher A.C. Methods of multivariate analysis. Wiley. 2002.
- Иберла К. Факторный анализ. М.: Статистика. 1980.
- 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.
- Машин В.А. Факторный анализ спектра сердечного ритма. Биофизика. 2011. Том 56, Вып. 2.

✓ Какой тип корреляции был использован при анализе главных компонент? Использование **корреляции Пирсона** для порядковых данных является **не корректным**. Корреляция Пирсона используется для интервальных данных. Для порядковых данных необходимо использовать **ранговые коэффициенты корреляции** (ранговый коэффициент корреляции Спирмена или ранговый коэффициент корреляции Кендалла).

- ✓ Матрица факторных нагрузок для РСА отсутствует.
- ✓ Матрица собственных значений для РСА отсутствует.
- ✓ Матрица значений общности для РСА отсутствует.

на 9 отобранных факторов приходилось **58%** дисперсии данных.

Это означает, что **42%** дисперсии в ответах респондентов на вопросы анкеты по культуре безопасности обусловлены другими факторами (фактором).

➤ В представленном виде невозможно использовать результаты РСА.

Например, случаи, когда многие респонденты не ответили на утверждение анкеты (т.е., **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.

Модели анализа главных компонент имеют ряд недостатков или ограничений. Одно из них состоит в том, что совершенно не очевидно, как поступать в ситуации, когда мы имеем дело с неполным набором данных (**пропущенные данные**). На сегодняшний день, неполные данные либо исключаются, либо заполняются с использованием различных методов интерполяции. Однако, такие подходы не действуют, когда значимая часть матрицы измерений неизвестна.

See more

- Haifeng Chen. Principal Component Analysis With Missing Data and Outliers
- ✓ Какой метод использовался для заполнения пропущенных данных при выполнении PCA? Какая часть матрицы измерений была неизвестна?
Замечание: Если переменные не отвечают требованию нормального распределения, мы не можем использовать средние оценки для пропущенных данных.

Мы провели отдельно PCA для 60 пунктов, которые были сохранены в анкете, и обнаружили, что семь из девяти исходных факторы выделяются в качестве отдельных факторов.

- ✓ Проверка допущения о многомерном нормальном распределении исходных переменных отсутствует;
- ✓ Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных отсутствует;
- ✓ Проверка целесообразности применения факторной модели анализа к исходным данным отсутствует;
- ✓ Выполнение процедуры отбора общих факторов отсутствует;
- ✓ Матрица факторных нагрузок для PCA отсутствует;
- ✓ Матрица собственных значений для PCA отсутствует;
- ✓ Матрица значений общности для PCA отсутствует;
- ✓ Суммарный процент дисперсии 7 факторов отсутствует.
- ✓ Использование матрицы корреляций Пирсона для порядковых данных является не корректным.

➤ В представленном виде невозможно использовать результаты PCA.

4.1. Исследовательский факторный анализ

(Метод главных осей - Principal Axis Factoring, PAF)

- ✓ Проверка допущения о многомерном нормальном распределении исходных переменных отсутствует;
 - ✓ Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных отсутствует;
 - ✓ Выбор метода вращения главных осей отсутствует;
 - ✓ Проверка целесообразности применения факторной модели анализа к исходным данным отсутствует;
 - ✓ Выполнение процедуры отбора общих факторов отсутствует;
 - ✓ Оценка значимости полученного факторного решения с помощью статистических критериев и принципа "простой структуры" отсутствует;
 - ✓ Суммарный процент дисперсии выделенных факторов отсутствует.
-
- ✓ Использование матрицы корреляций Пирсона для порядковых данных является не корректным.

➤ В представленном виде невозможно использовать результаты PAF.

4.2. Анализ надежности

Мы рассчитали альфа-коэффициент Кронбаха для каждого из факторов и субфакторов, выявленных INPO, используя сокращенных до 60 пунктов опросник.

Рекомендуется использовать порядковый коэффициент альфа и порядковый коэффициент тета для коррекции отрицательного смещения альфа-коэффициента Кронбаха, в связи с данными ответов по шкале Лайкерта.

См. дополнительно:

- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta For Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, 6, 21-29.

4.4. Анализ внутригрупповой надежности

Два типа коэффициента внутригрупповой корреляции (ICC) используются для определения внутригрупповой надежности: ICC(1) измеряет надежность среди индивидов в группе, и ICC(2) измеряет надежность средних значений групп.

- ✓ Для использования ICC(1) и ICC(2) должны быть рассмотрены следующие основные предположения:
 - Равные (похожие) дисперсии
 - Многомерное нормальное распределение данных (приблизительно)
 - **Интервальный** характер данных (шкала Лайкерта – **порядковая** шкала)
 - Независимость выборок (т.е., результаты наблюдений для одного человека должны быть независимы от результатов наблюдений для другого человека, которые проходил анкетирование)

➤ В представленном виде невозможно использовать результаты ICC(1) и ICC(2).

См. дополнительно:

- Daniel Stahl Introduction to measurement and scale development. Department of Biostatistics & Computing. King's College London. 2007.

Таблица 6. Взаимная корреляция между общим индексом культуры безопасности и отдельными факторами культуры безопасности

➤ Невозможно использовать результаты факторного анализа (см. выше).

5.2. Конкурентная валидность данных опроса и показателей эффективности деятельности NRC²

Статистикой, используемой для проверки взаимосвязи между культурой безопасности и показателями эксплуатационной безопасности, был коэффициент корреляции Пирсона. **Коэффициент корреляции Пирсона** является **параметрической** статистикой, которая проверяет линейную зависимость между двумя переменными.

Корреляция Пирсона предназначена для использования, когда обе переменные измерены либо **интервальной** шкалой, либо шкалой **отношений**, и каждая переменная **нормально распределена**.

² Nuclear Regulatory Commission - Комиссия по ядерному регулированию

- ✓ Переменные культуры безопасности измерены **порядковой** шкалой. Для порядковых данных необходимо использовать **ранговые коэффициенты корреляции** (ранговый коэффициент корреляции Спирмена или ранговый коэффициент корреляции Кендалла).

Факторы культуры безопасности опросника также удовлетворяют предположению о нормальности распределения, что было установлено с помощью критерия Шапиро-Уилка (Shapiro-Wilk) для проверки **нормальности** распределений переменных.

- ✓ Результаты критерия Шапиро-Уилка для нормальности распределения факторов культуры безопасности опросника отсутствуют. (Ненормальные распределения являются Нормой в реальном мире.)

См. дополнительно:

- Thomas Puzdek. Non-Normal Distributions in the Real World. 2000.

Показатели эффективности деятельности АЭС, используемые NRC, **не** демонстрировали **нормальное** распределение и не прошли проверку по критерию Шапиро-Уилка на нормальность.

Ранговые коэффициенты корреляции Спирмена и Кендалла являются **непараметрическими** статистиками и не требуют выполнения предположения о **нормальности** распределения переменных.

- ✓ Если мы имеем данные, которые ненормально распределены, мы не можем использовать любые статистики, которые предполагают нормальное распределение переменных (например, ANOVA, корреляция Пирсона, критерий Стьюдента t). Если наши данные нормально не распределены, часто возможно нормализовать распределение переменных через их преобразование (нормализация данных).

Мы рассчитали корреляции между результатами ответов на опросник культуры безопасности и показателями безопасности эксплуатации АЭС используя коэффициент корреляции Пирсона и ранговый коэффициент корреляции **Кендалла**.

На ранговый коэффициент корреляции Кендалла могут оказывать существенное влияние связи в данных (например, несколько АЭС имеют одинаковое количество незапланированных срабатываний аварийной защиты), что ведет к ложным низким коэффициентам корреляции.

- ✓ Рекомендуется использовать коэффициент корреляции Спирмена при наличии одинаковых рангов в переменных. В этом случае вводится поправка на одинаковые ранги в расчет коэффициента.

См. дополнительно:

- Суходольский Г.В. Основы математической статистики для психологов. Изд.: Санкт-Петербургский университет, 1998.

- ✓ Значения рангового коэффициент корреляции Кендалла отсутствуют.

➤ В представленном виде невозможно использовать результаты корреляционного анализа между результатами ответов на опросник культуры безопасности и показателями безопасности эксплуатации АЭС для анализа конкурентной валидности.

См. дополнительно:

- SPSS for Windows. Version 19.0: A Basic Tutorial. Chapter Seven: Correlation and Regression. 2010 (http://ssric.org/files/spss_v19.pdf)

С учетом результатов предыдущих исследований, мы ожидали слабые или умеренные уровни влияния с коэффициентами корреляции от **.20 до .30**.

Коэффициент детерминации (r^2) равен квадрату коэффициента корреляции Пирсона r . Коэффициент детерминации, с точки зрения регрессионного анализа (линейная модель), это доля дисперсии зависимой переменной (показатели безопасности), объясняемая независимой переменной (факторы культуры безопасности, согласно результатам обследования).

Если коэффициенты корреляции лежат в диапазоне от **.20 до .30**, тогда коэффициент детерминации r^2 находится в диапазоне от **.04 до .09** (от **4% до 9%** дисперсии показателей безопасности эксплуатации АЭС могут быть объяснены влиянием факторов культуры безопасности, полученных в ходе обследования). Это **очень слабая связь** между переменными, согласно шкале Чеддока (Cheddoka) (r^2 от .10 до .30 – слабая взаимосвязь).

Таблица 7 представляет корреляцию Пирсона для оценки, конкурентной валидности, между культурой безопасности, индикаторами эффективности деятельности NRC, результатами инспекций, а также Матрицей действий по результатам ROP³ in 2010. **Table 8** представляет коэффициенты корреляции между культурой безопасности, сообщениями персонала по вопросам безопасности для NRC, компоненты и факторы широкого влияния⁴ в рамках ROP за 2010.

Корреляции, которые отмечены в таблицах как статистически значимые:

Таблица 7: $r = -.25 \div -.46$ $r^2 = .06 \div 0.21$ (**очень слабая – слабая связь**)

Таблица 8: $r = -.25 \div -.48$ $r^2 = .06 \div 0.23$ (**очень слабая – слабая связь**)

➤ В представленном виде невозможно использовать результаты корреляционного анализа между результатами опросника культуры безопасности и показателями безопасности эксплуатации АЭС для анализа конкурентной валидности.

(В любом случае, результаты предполагают очень слабую или незначительную связь между факторами опросника культуры безопасности и ключевыми индикаторами показателей безопасности АЭС.)

5.3. Исследование прогностической валидности опросника с помощью показателей безопасности деятельности NRC

Таблица 9 Корреляция между Культурой безопасности и компонентами и факторами широкого влияния в рамках ROP за 2011

Таблица 10 Корреляция между Культурой безопасности, Матрицей действий по результатам ROP, результатами инспекций, SCCI⁵, и сообщениями персонала в NRC за 2011

Таблица 9: $r = -.25 \div -.30$ $r^2 = .06 \div 0.09$ (**очень слабая связь**)

Таблица 10: $r = -.26 \div -.48$ $r^2 = .07 \div 0.23$ (**очень слабая – слабая связь**)

³ Reactor Oversight Process - процесс реакторного надзора

⁴ Например, человеческий фактор, культура справедливых отношений, эффективность выявления и решения проблем безопасности.

⁵ Substantive Cross-Cutting Issue – Основные вопросы широкого влияния

Замечание к Таблицам 9 и 10

Статистическая значимость коэффициента корреляции Пирсона (r) = 0.048024 для $p = 0.01$ ($N = 2876$) и $r = 0.036551$ для $p = 0.05$ ($N = 2876$). Это результат большого размера выборки (N).

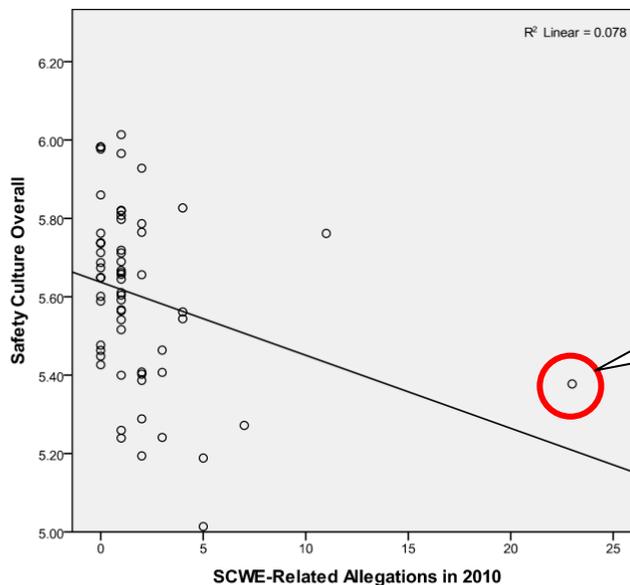


Figure 6 Scatterplot of Safety Culture and SCWE-Related Allegations in 2010

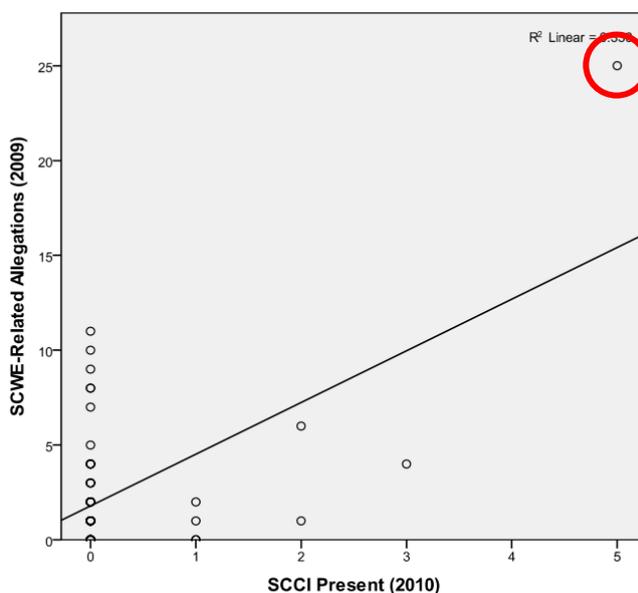


Figure 9 Scatterplot of SCWE-related Allegations in 2009 and SCCIs in 2010

- ✓ Анализ выбросов отсутствует.
- ✓ Анализ причин выбросов отсутствует.
- ✓ Использование матрицы корреляций Пирсона для порядковых данных является не корректным.

➤ В представленном виде невозможно использовать результаты корреляционного анализа между результатами опросника культуры безопасности и показателями безопасности эксплуатации АЭС для исследования прогностической валидности.

(В любом случае, эти результаты предполагают очень слабую или незначительную связь между переменными.)

Основные выводы

Согласно Культуре статистического анализа крайне важно:

- Соответствующее использование статистических методов в исследовании,
- Учитывать ограничения статистических методов и основные предположения.
- Рассчитывать размер выборки для проведения опроса, учитывая доверительный интервал, доверительную вероятность и размер генеральной совокупности.
- Учитывать тип данных при выборе статистических методов.
- Понимать, что использование корреляции Пирсона для порядковых данных является некорректным.
- ✓ При использовании анализа главных компонент (PCA) обязательно рассматривать:
 - Проверку допущения о многомерном нормальном распределении исходных переменных;
 - Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных;
 - Проверку целесообразности применения факторной модели анализа к исходным данным;
 - Выполнение процедуры отбора общих факторов.
- Для порядковых данных использовать ранговые коэффициенты корреляции при анализе главных компонент (PCA).
- ✓ Для использования метода главных осей (PAF) обязательно рассматривать:

- Проверку допущения о многомерном нормальном распределении исходных переменных;
 - Выбор типа матрицы парных коэффициентов корреляции в качестве исходных данных;
 - Выбор метода вращения главных осей;
 - Проверку целесообразности применения факторной модели анализа к исходным данным;
 - Выполнение процедуры отбора общих факторов;
 - Оценку значимости полученного факторного решения с помощью статистических критериев и принципа "простой структуры".
- Для порядковых данных использовать ранговые коэффициенты корреляции при анализе методом главных осей (РАФ).
- Для использования коэффициентов внутригрупповой корреляции ИСС(1) и ИСС(2) обязательно рассматривать:
- Равность дисперсий
 - Проверку допущения о многомерном нормальном распределении данных (приблизительное)
 - Интервальный характер данных (шкала Лайкерта – порядковая шкала)
 - Независимость выборок (т.е., результаты наблюдений для одного человека должны быть независимы от результатов наблюдений для другого человека, которые проходил анкетирование)
- Для анализа коэффициентов корреляции Пирсона проверить данные на выбросы.

Статистический анализ результатов исследования культуры безопасности с помощью опросника INPO, представленный в статье, делает невозможным получение достоверных выводов.