Mashin V.A. Comments on the Technical Paper

Independent Evaluation of INPO's Nuclear Safety Culture Survey and Construct Validation Study. Stephanie Morrow, 2012.

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

Mashin V.A.
The Main Specialist of Central Institute of Continuing Education and Training, Dr. of psychology.
249031 Russia, Obninsk, Kaluga region, CICE&T, 21, Kurchatov str. E-mail: mashin-va@mail.ru

2.4. Relationships between Safety Culture and Safety Performance

The results from the meta-analytic studies found consistent evidence of a statistically significant linear relationship between safety culture and accidents/injuries, ranging from a correlation of **-.22** to **-.39** ($p < .05$).
Using Cohen's labels, the relationship between safety culture and safety performance appears to be a **medium** effect.

According to Achim Bühl; Peter Zöfel (2001):

$0.0 < r \leq 0.2$	very small effect
$0.2 < r \leq 0.5$	**small effect**
$0.5 < r \leq 0.7$	medium effect
$0.7 < r \leq 0.9$	large effect
$0.9 < r \leq 1.0$	very large effect

See more:

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

3.2. Survey Administration

(Determining of Sample Size)

The Central Limit Theorem posits that as the size of a random sample increases the sample mean approaches the population mean (Myers & Well, 2003). When the sample size is greater than **30** it tends to approximate a normal distribution and there are negligible differences between the sample mean and population mean. Thus, a sample size of **30** is commonly used as a minimum threshold to establish that the sample is a good estimate for the population.

The Central Limit Theorem not use for determine of Sample Size in the study of population. And $N \geq 30$ for sample is an ***arbitrary*** number. The statement of the Central Limit Theorem does not have a statement about $N \geq 30$ and '*approximate a normal distribution*' and '*negligible differences between the sample mean and population mean*'.

✓ We must account for determine of Sample Size:
- Confidence Level
- Confidence Interval
- Population

Sample size allows us to set the Confidence Level and Confidence Interval for the average values for the population.

See more:
Sample Size Calculator (Creative Research Systems)
http://www.surveysystem.com/sscalc.htm

Example:
- Confidence Level = 0.95
- Confidence Interval = 5
- Population = 600

Sample size needed: = **234**

➢ Sample size does not meet the statistical requirements

Look also:
- 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. Survey Item Development

(The data type)

Survey participants were asked to rate their degree of agreement with each statement using a 7-point Likert scale ranging from strongly disagree to strongly agree. The scoring system from 1-7 enables the researcher to ascribe a quantitative value to the respondent's qualitative assessment. The data are traditionally used as an approximation of an **interval scale**, meaning that the response options are inferred to be relatively equidistant from each other on a continuum.

Likert scale is **ordinal scale** on which data is shown simply in order of magnitude since there is no standard of measurement of differences:

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

Interval scale is a scale of measurement of data according to which the differences between values can be quantified in absolute but not relative terms and for which any zero is merely arbitrary. Interval scale have units with measured pace (degrees, second, etc.). The measured object is assigned a number equal to the number of units it contains:

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

➢ It is not correct use Likert scale as interval scale.

See examples from:

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

4.1. Exploratory Factor Analysis

(Principal Components Analysis - PCA)

INPO's approach to assessing the dimensionality of safety culture, based on the items included in the safety culture survey, was to perform a type of exploratory factor analysis called **principal components analysis (PCA)**.

✓ For use PCA have not been considered:
- the multi-variate normal assumption,
- the type of correlation coefficients matrix as input data,
- factorability of the intercorrelation matrix,
- criteria for determining the number of factors.

See more

- Rencher A.C. Methods of multivariate analysis. Wiley. 2002.
- Iberla K. The factorial analysis. 1980.
- Kim J. O., Muller C. W., Klekka W. R. et al. Factor, discriminate and cluster analysis. 1989.
- Mashin V. A. Factor analysis of the heart rate spectrum. Biofizika. 2011 Mar-Apr; 56(2):328-41.

✓ What type the correlation is used in principal components analysis?
The use of **Pearson correlation** for ordinal data is **not correct**. Pearson correlation is used for interval data. For ordinal data need to use **Rank correlation coefficients** (Spearman's rank correlation coefficient or Kendall tau rank correlation coefficient).
The use of the covariance does not change anything. Between covariance (*cov*) and Pearson correlation (*r*) there is a simple relationship:
$r(x,y) = cov(X,Y)/[\sigma(x)\times\sigma(y)]$; where: $\sigma(x)$, $\sigma(y)$ - standard deviation x and y.

✓ The factors loadings matrix of PCA is missing.
✓ The Eigenvalues matrix of PCA is missing.
✓ The Communalities matrix of PCA is missing.

the nine factors accounted for **58%** of the variance in the data.

It is mean that **42%** of the variance in the answers of the respondents on Safety Culture survey is related with other factors (factor).

➢ <u>In the present form is not possible to use the results of PCA.</u>

For instance, cases where many respondents did not respond to an item (i.e., **missing data**), or responded by choosing the “don’t know” option may indicate that respondents were confused by the item or did not find it applicable to their work.

PCA models have several **shortcomings**. One is that it is not obvious how to deal properly with **incomplete data set**, in which some of the points are missing. Currently the incomplete points are either discarded or completed using a variety of interpolation methods. However, such approaches are no longer valid when a significant portion of the measurement matrix is unknown.

See more

- Haifeng Chen. Principal Component Analysis With Missing Data and Outliers

✓ What the method are used to solve this drawback of standard PCA? What is portion of the measurement matrix is unknown?
Note: If the variables do not meet the requirement of a normal distribution we cannot use mean score for missing data.

We conducted an independent PCA of the 60 items that were retained in the survey and found that seven of the original nine factors emerged as distinct factors.

✓ The multi-variate normal assumption is missing.
✓ Select the type of correlation coefficients matrix as input data is missing.
✓ Factorability of the intercorrelation matrix is missing.
✓ Criteria for determining the number of factors is missing.
✓ The Eigenvalues matrix of PCA is missing.
✓ The Communalities matrix of PCA is missing.
✓ The factors loadings matrix of PCA is missing.
✓ The cumulative percent of variance of seven factors is missing.

✓ The use of Pearson correlation matrix for ordinal data is not correct.

➢ <u>In the present form is not possible to use the results of PCA.</u>

4.1. Exploratory Factor Analysis

(Principal Axis Factoring - PAF)

✓ The multi-variate normal assumption is missing.
✓ Select the type of correlation coefficients matrix as input data is missing.
✓ Select the rotation method of Principal Axis is missing.
✓ Factorability of the intercorrelation matrix is missing.

- ✓ Criteria for determining the number of factors is missing.
- ✓ The validity of the factor analysis model.
- ✓ The Eigenvalues matrix is missing.
- ✓ The Communalities matrix is missing.
- ✓ The factors loadings matrix is missing.
- ✓ The cumulative percent of variance of extracting factors is missing.

- ✓ The use of Pearson correlation for ordinal data is not correct.

- ➢ <u>In the present form is not possible to use the results of PAF.</u>

4.2. Reliability Analysis

We calculated Cronbach's coefficient alpha for each of the factors and sub-factors identified by INPO using the reduced 60 item survey.

It is recommended to use the ordinal coefficient alpha or ordinal coefficient theta to correct for the negative bias in coefficient alpha due to Likert response data.

See more:

- 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. Within-Group Reliability Analysis

Two types of ICCs are relevant for determining within-group reliability: ICC(1) measures reliability among individuals in a group, and ICC(2) measures the reliability of the group mean.

- ✓ For use ICC(1) and ICC(2) have not been considered the main assumptions:
  - Equal (similar) variances
  - (Multi-) Normal distributed data (roughly)
  - **Interval** level data (Likert scale is **ordinal** scale)
  - Independence (e.g. observations for one person should be independent of observations for any other person who did get the test)

➢ In the present form is not possible to use the results of ICC(1) and ICC(2).

See more:

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

Table 6 Intercorrelations among Safety Culture Overall and Safety Culture Factors

➢ It is not possible to use results of Factor analysis (look before).

5.2. Concurrent Validity of the Survey Data with NRC Performance Metrics

The statistic used to test for a relationship between safety culture and safety performance was Pearson's product-moment correlation coefficient. **Pearson's correlation** is a **parametric statistic** that tests the linear relationship between two variables.

The Pearson correlation is intended to be used when both variables are measured at either the **interval** or **ratio** level, and each variable is **normally** distributed.

✓ The variables of Safety Culture are measured at the **ordinal scale**. For ordinal data need to use **Rank correlation coefficients** (Spearman's rank correlation coefficient and Kendall tau rank correlation coefficient).

The survey factors also meet assumptions of normality, as determined by the Shapiro-Wilk test of **normality**.

✓ The results Shapiro-Wilk test of normality are missed.
(Non-Normal Distributions is Norma in the Real World.)

See more:

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

The NRC performance metrics do **not** demonstrate **normal** distributions and fail to meet the Shapiro-Wilk test of normality.

Spearman's rank correlation coefficient and Kendall tau rank correlation coefficient are **nonparametric** statistics and do not require the assumption of **normality**.

- ✓ If our data are not normally distributed we cannot use any of the tests that assume that it is (e.g. ANOVA, Pearson correlation, t test, regression analysis). If our data are not normally distributed it is often possible to normalise it by transforming it.

We calculated correlations between the safety culture survey results and safety performance variables using Pearson's correlation coefficient and the **Kendall** Tau correlation coefficient.

The Kendall Tau statistic can be significantly influenced by ties in the data (e.g., multiple sites have the same number of unplanned scrams), resulting in spuriously small correlation coefficients.

- ✓ It is recommended to use Spearman's rank correlation coefficient. Using Spearman's rank correlation coefficient need to be amended on identical ranks in calculation if variables have groups with identical ranks.

See more:

- Sukhodolskiy G.V. Fundamentals of Mathematical Statistics for Psychologists. 1998.

- ✓ Kendall Tau correlation coefficients are missed.

- ➢ In the present form is not possible to use the results of correlations analysis between the safety culture survey results and safety performance variables for Concurrent Validity analysis.

See more

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

Given the results of previous studies, we expected small to medium effect sizes with correlation coefficients of **.20 to .30**.

The coefficient of determination (r^2), is the square of the Pearson correlation coefficient *r*. The coefficient of determination, with respect to regression analysis (the linear model), is the proportion of the dependent variable variance (safety performance) explained by the independent variable variance (safety culture survey).

If correlation coefficients of **.20 to .30** then the coefficient of determination r^2 = **.04 to .09** (**4% to 9%** the amount of safety performance variation that can be explained by safety culture survey variation). It is **very weak relationship between the variables** on a scale Cheddoka (r^2 = .10 to .30 - weak relationship).

Table 7 presents the concurrent Pearson's correlations between safety culture, NRC performance indicators, inspection findings, and the ROP Action Matrix in 2010. **Table 8** presents the correlations between safety culture, NRC allegations, and ROP cross-cutting areas and components in 2010.

Correlations that are statistically significant:

Table 7: -.25 to -.46 r^2 = .06 to 0.21 (**very weak - weak relationship**)

Table 8: -.25 to -.48 r^2 = .06 to 0.23 (**very weak - weak relationship**)

- <u>In the present form is not possible to use the results of correlations analysis between the safety culture survey results and safety performance variables for Concurrent Validity analysis.</u>

  (In any case, the results suggest a very weak or weak relationship between the factors/sub-factors of safety culture survey and the key performance indicators.)

5.3. Exploratory Predictive Validity of the Survey with NRC Performance Metrics

Table 9 Correlations between Safety Culture and 2011 ROP Cross-Cutting Areas and Components

Table 10 Correlations between Safety Culture and 2011 ROP Action Matrix, Inspection Findings, SCCIs, and Allegations

Table 9: -.25 to -.30 r^2 = .06 to 0.09 (**very weak relationship, very few significant** Pearson correlation coefficients)

Table 10: -.26 to -.48 r^2 = .07 to 0.23 (**very weak - weak relationship**)

Remark for Tables 9 and 10

Statistically significant Pearson Correlation (r) = 0.048024 for p = 0.01 (N = 2876) and r = 0.036551 for p = 0.05 (N = 2876). This is the result of large sample size (N).

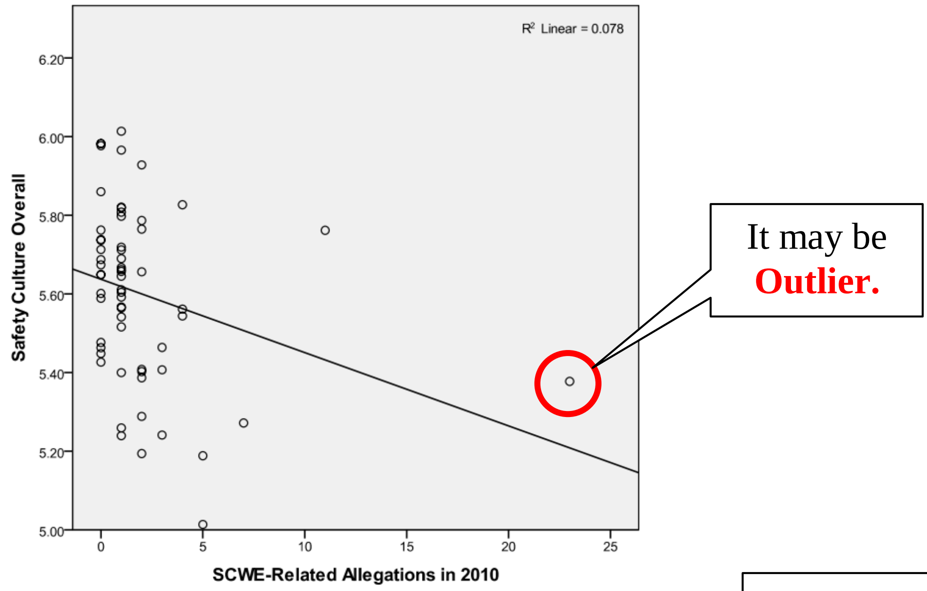

Figure 6 Scatterplot of Safety Culture and SCWE-Related Allegations in 2010

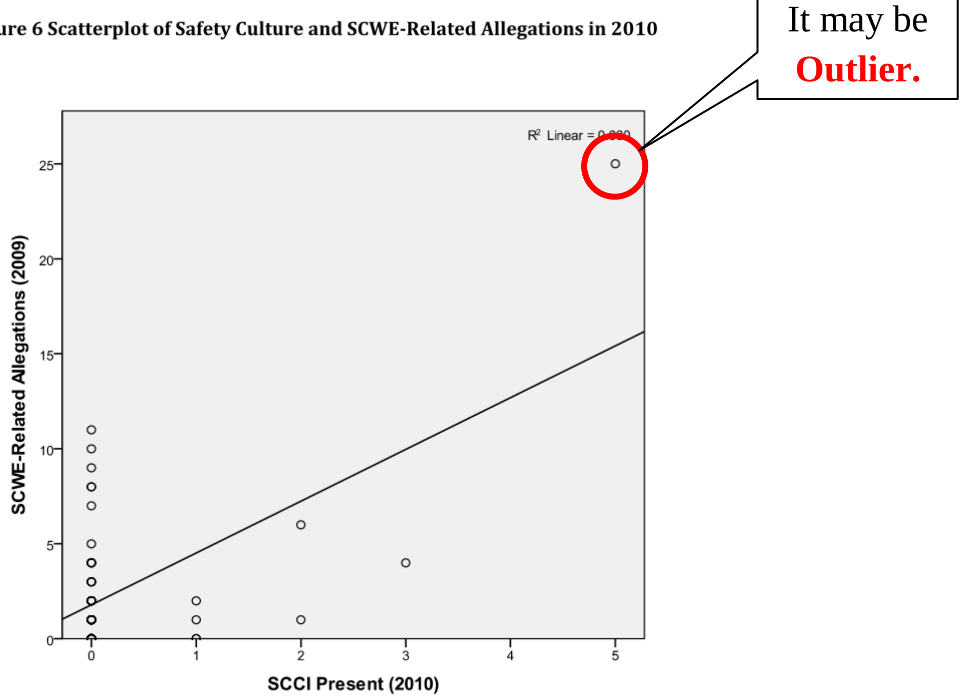

Figure 9 Scatterplot of SCWE-related Allegations in 2009 and SCCIs in 2010

- ✓ Outlier analysis is missed.
- ✓ Outlier causes analysis is missed.
- ✓ The use of Pearson correlation for ordinal data is not correct.

- In the present form is not possible to use the results of correlations analysis between the safety culture survey results and safety performance variables for Exploratory Predictive Validity.

(In any case, the results suggest a very weak or weak relationship between the variables.)

The main Conclusions

According to Statistical Culture it is very important:

- Appropriate use of the statistical methods in the study.
- Account for limitations of the statistical methods and the main assumptions.
- Calculation of Sample Size account for Confidence Level, Confidence Interval, Population.
- Account for the data type when choosing the statistical methods.
- Understand what to use of Pearson correlation for ordinal data of Likert scale is not correct.
- ✓ For the use of Principal Components Analysis (PCA) should be considered:
  - the multi-variate normal assumption,
  - the type of correlation coefficients matrix as input data,
  - factorability of the intercorrelation matrix,
  - criteria for determining the number of factors.
- With ordinal data should be used Rank correlation coefficients for PCA.
- ✓ For the use of Principal Axis Factoring (PAF) should be considered:
  - the multi-variate normal assumption,
  - the type of correlation coefficients matrix as input data,
  - the rotation method of Principal Axis,
  - factorability of the intercorrelation matrix,
  - criteria for determining the number of factors.
  - the validity of the factor analysis model.
- With ordinal data should be used Rank correlation coefficients for PAF.
- For the use of ICC(1) and ICC(2) should be considered:

- Equal (similar) variances
- (Multi-) Normal distributed data (roughly)
- Interval level data (Likert scale is ordinal scale)
- Independence (e.g. observations for one person should be independent of observations for any other person who did get the test)

➢ For the analysis of Pearson correlation coefficients should be considered data on outliers.

Statistical analysis of the safety culture survey presented in the Technical Paper makes it impossible to draw valid findings.