

Research Article

About Statistical Analysis of Qualitative Survey Data

Stefan Loehnert

IBM Beta Test Side Data Migration Center, 71083 Herrenberg, Germany

Correspondence should be addressed to Stefan Loehnert, stefan_loehnert@de.ibm.com

Received 4 August 2009; Accepted 15 April 2010

Academic Editor: Suk joo Bae

Copyright © 2010 Stefan Loehnert. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gathered data is frequently not in a numerical form allowing immediate appliance of the quantitative mathematical-statistical methods. In this paper are some basic aspects examining how quantitative-based statistical methodology can be utilized in the analysis of qualitative data sets. The transformation of qualitative data into numeric values is considered as the entrance point to quantitative analysis. Concurrently related publications and impacts of scale transformations are discussed. Subsequently, it is shown how correlation coefficients are usable in conjunction with data aggregation constrains to construct relationship modelling matrices. For illustration, a case study is referenced at which ordinal type ordered qualitative survey answers are allocated to process defining procedures as aggregation levels. Finally options about measuring the adherence of the gathered empirical data to such kind of derived aggregation models are introduced and a statistically based reliability check approach to evaluate the reliability of the chosen model specification is outlined.

1. Introduction

In this paper some aspects are discussed how data of qualitative category type, often gathered via questionnaires and surveys, can be transformed into appropriate numerical values to enable the full spectrum of quantitative mathematical-statistical analysis methodology. Therefore the impacts of the chosen valuation-transformation from ordinal scales to interval scales and their relations to statistical and measurement modelling are studied. This is applied to demonstrate ways to measure adherence of quantitative data representation to qualitative aggregation assessments-based on statistical modelling. Finally an approach to evaluate such adherence models is introduced. Concurrent a brief epitome of related publications is given and examples from a case study are referenced.

Gathering data is referencing a data typology of two basic modes of inquiry consequently associated with “qualitative” and “quantitative” survey results. Thereby “quantitative” is looked at to be a response given directly as a numeric value and “qualitative” is a nonnumeric answer. This differentiation has its roots within the social sciences and research. A brief comparison of this typology is given in [1, 2]. A refinement by adding the predicates “objective”

and “subjective” is introduced in [3]. An elaboration of the method usage in social science and psychology is presented in [4]. A precis on the qualitative type can be found in [5] and for the quantitative type in [6]. A comprehensive book about the qualitative methodology in social science and research is [7]. Since both of these methodic approaches have advantages on their own it is an ongoing effort to bridge the gap between, to merge, or to integrate them. Following [8], the conversion or transformation from qualitative data into quantitative data is called “quantizing” and the converse from quantitative to qualitative is named “qualitizing”. The research on mixed method designs evolved within the last decade starting with analysis of a very basic approach like using sample counts as quantitative base, a strict differentiation of applying quantitative methods to quantitative data and qualitative methods to qualitative data, and a significant loose of context information if qualitative data (e.g., verbal or visual data) are converted into a numerically representation with a single meaning only [9].

A well-known model in social science is “triangulation” which is applying both methodic approaches independently and having finally a combined interpretation result. The main mathematical-statistical method applied thereby is cluster-analysis [10]. Model types with gradual differences

in methodic approaches from classical statistical hypothesis testing to complex triangulation modelling are collected in [11]. Recently, it is recognized that mixed methods designs can provide pragmatic advantages in exploring complex research questions. However, the analytic process of analyzing, coding, and integrating unstructured with structured data by applying quantizing qualitative data can be a complex, time consuming, and expensive process. In [12], Driscoll et al. are presenting an example with simple statistical measures associated to strictly different response categories whereby the sample size issue at quantizing is also sketched.

A way of linking qualitative and quantitative results mathematically can be found in [13]. There are fuzzy logic-based transformations examined to gain insights from one aspect type over the other. Also in mathematical modeling, qualitative and quantitative concepts are utilized. In terms of decision theory [14], Gascon examined properties and constraints to timelines with LTL (linear temporal logic) categorizing qualitative as likewise nondeterministic structural, for example, cyclic, and quantitative as a numerically expressible identity relation. The object of special interest thereby is a symbolic representation of a \mathbf{Z} -valuation with \mathbf{Z} denoting the set of integers. A symbolic representation defines an equivalence relation between \mathbf{Z} -valuations and contains all the relevant information to evaluate constraints. This might be interpreted as a hint that quantizing qualitative surveys may not necessarily reduce the information content in an inappropriate manner if a valuation similar to a \mathbf{Z} -valuation is utilized. In [15] Herzberg explores the relationship between propositional model theory and social decision making via premise-based procedures. Condensed it is exposed that certain ultrafilters, which in the context of social choice are decisive coalitions, are in a one-to-one correspondence to certain kinds of judgment aggregation functions constructed as ultra-products. A special result is a “Impossibility theorem for finite electorates” on judgment aggregation functions, that is, if the population is endowed with some measure-theoretic or topological structure, there exists a single overall consistent aggregation.

2. Interlock Qualitative and Quantitative Concepts

2.1. From Quantitative Results to Qualitative Insights. Fuzzy logic-based transformations are not the only examined options to qualitzing in literature. The transformation from quantitative measures into qualitative assessments of software systems via judgment functions is studied in [16]. Based on Dempster-Shafer belief functions, certain objects from the realm of the mathematical theory of evidence [17], Kłopotek and Wierzchon. utilized exemplified decision tables as a (probability) measure of diversity in relational data bases. The authors viewed the Dempster-Shafer belief functions as a subjective uncertainty measure, a kind of generalization of Bayesian theory of subjective probability and showed a correspondence to the join operator of the relational database theory. This rough set-based representation

TABLE 1: Effectiveness rate example.

	Self-assessment to initial	Initial to follow-up
Counting change rate	5%	12,5%
Variance change rate	14,3%	3,7%

of belief function operators led then to a nonquantitative interpretation. As a more direct approach the net balance statistic as the percentage of respondents replying “up” less the percentage replying “down” is utilized in [18] as a qualitative yardstick to indicate the direction (up, same or down) and size (small or large) of the year-on-year percentage change of corresponding quantitative data of a particular activity.

The following real life-based example demonstrates how misleading pure counting-based tendency interpretation might be and how important a valid choice of parametrization appears to be especially if an evolution over time has to be considered.

Example 1 (A Misleading Interpretation of Pure Counts). The situation and the case study-based on the following: projects (m) are requested to answer to an ordinal scaled survey about alignment and adherence to a specified procedural-based process framework in a self-assessment. Then the ($n = 104$) survey questions are worked through with a project external reviewer in an “initial review”. Based on these review results improvement recommendations are given to the project team. After a certain period of time a follow-up review was performed. So three samples available: self-assessment, initial review and follow-up sample. In case of such timeline depending data gathering the cumulated overall counts according to the scale values are useful to calculate approximation slopes and allow some insight about how the overall projects behavior evolves. Now we take a look at the pure counts of changes from self-assessment to initial review which turned out to be 5% of total count and from the initial review to the follow-up with 12,5% changed. Misleading is now the interpretation that the effect of the follow-up is greater than the initial review effect. Obviously the follow-up is not independent of the initial review since recommendations are given previously from initial review. A better effectiveness comparison is provided through the usage of statistically relevant expressions like the variance. For the self-assessment the answer variance was 6,3(%), for the initial review 5,4(%) and for the follow-up 5,2(%). This leads to the relative effectiveness rates shown in Table 1.

A variance-expression is the one-dimensional parameter of choice for such an effectiveness rating since it is a deviation measure on the examined subject-matter. The mean (or median or mode) values of alignment are not as applicable as the variances since they are too subjective at the self-assessment, and with high probability the follow-up means are expected to increase because of the outlined improvement recommendations given at the initial review.

Thereby, the empirical unbiased question-variance σ_Q^2 is calculated from the survey results with x_{ij} as the j th answer to question i and the according expected single question means μ_i , that is,

$$\sigma_Q^2 = \frac{1}{n-1} \sum_{i=1}^n (\mu_i - \mu)^2 \quad \text{with } \mu = \frac{1}{n} \sum_{i=1}^n \mu_i, \mu_i = \frac{1}{m} \sum_{j=1}^m x_{ij}. \quad (1)$$

In contrast to the one-dimensional full sample mean

$$\mu = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m x_{ij} \quad (2)$$

which is identical to the summing of the single question means μ_i , σ_Q^2 is not identical to the unbiased empirical full sample variance

$$\sigma_S^2 = \frac{1}{n \cdot m - 1} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \mu)^2. \quad (3)$$

Also it is not identical to the expected answer mean variance

$$\hat{\sigma}_*^2 := E(\sigma_*^2) = \frac{1}{n} \sum_{i=1}^n \sigma_i^2 = \frac{1}{n \cdot (m-1)} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \mu)^2 \quad (4)$$

where by the answer variance at the i th question is

$$\sigma_i^2 = \frac{1}{m-1} \sum_{j=1}^m (x_{ij} - \mu_i)^2. \quad (5)$$

It is a qualitative decision to use σ_Q^2 triggered by the intention to gain insights of the overall answer behavior. The full sample variance σ_S^2 might be useful at analysis of single project answers, σ_*^2 in the context of question comparison and σ_i^2 for a detailed analysis of the specified single question. So σ_i^2 is useful to evaluate the applied compliance and valuation criteria or to determine a predefined review focus scope. In fact, to enable such a kind of statistical analysis it is needed to have the data available as, respectively, transformed into, an appropriate numerical coding.

2.2. Transforming Qualitative Data for Quantitative Analysis.

The research and appliance of quantitative methods to qualitative data has a long tradition. Due to [19] is the method of "Equal-Appearing Interval Scaling". Essentially this is to choose a representative statement (e.g., to create a survey) out of each group of statements formed from a set of statements related to an attitude using the median value of the single statements as grouping criteria. A single statement's median is thereby calculated from the "favourableness" on a given scale assigned to the statement towards the attitude by a group of judging evaluators. A link with an example can be found at [20] (Thurstone Scaling). Also the technique of correspondence analyses, for instance, goes back to research in the 40th of the last century for a compendium about the history see Gower [21]. Correspondence analysis is known also under different synonyms

like optimal scaling, reciprocal averaging, quantification method (Japan) or homogeneity analysis, and so forth [22] Young references to correspondence analysis and canonical decomposition (synonyms: parallel factor analysis or alternating least squares) as theoretical and methodological cornerstones for quantitative analysis of qualitative data. The great efficiency of applying principal component analysis at nominal scaling is shown in [23]. There is given a nice example of an analysis of business communication in the light of negotiation probability. The authors introduced a five-stage approach with transforming a qualitative categorization into a quantitative interpretation (material sourcing—transcription—unitization—categorization—nominal coding). The issues related to timeline reflecting longitudinal organization of data, exemplified in case of life history are of special interest in [24]. Thereby so-called Self-Organizing Maps (SOMs) are utilized. SOMs are a technique of data visualization accomplishing a reduction of data dimensions and displaying similarities. The authors consider SOMs as a nonlinear generalization of principal component analysis to deduce a quantitative encoding by applying life history clustering algorithm-based on the Euclidean distance (x, y n -dimensional vectors in Euclidian space)

$$d(x, y) = \sqrt{\sum_{s=1}^n (x_s - y_s)^2}. \quad (6)$$

Belief functions, to a certain degree a linkage between relation, modelling and factor analysis, are studied in [25]. The authors used them to generate numeric judgments with nonnumeric inputs in the development of approximate reasoning systems utilized as a practical interface between the users and a decision support system. Another way to apply probabilities to qualitative information is given by the so-called "Knowledge Tracking (KT)" methodology as described in [26]. Thereby the idea is to determine relations in qualitative data to get a conceptual transformation and to allocate transition probabilities accordingly. Thus the emerging cluster network sequences are captured with a numerical score ("goodness of fit score") which expresses how well a relational structure explains the data. Since such a listing of numerical scores can be ordered by the lower-less (\leq) relation KT is providing an ordinal scaling. Limitations of ordinal scaling at clustering of qualitative data from the perspective of phenomenological analysis are discussed in [27].

3. Scaling

It is a well-known fact that the parametrical statistical methods, for example, ANOVA (Analysis of Variance), need to have some kinds of standardization at the gathered data to enable the comparable usage and determination of relevant statistical parameters like mean, variance, correlation, and other distribution describing characteristics. A survey about conceptual data gathering strategies and context constrains can be found in [28]. One of the basics thereby is the underlying scale assigned to the gathered data.

The main types of numerically (real number) expressed scales are

- (i) nominal scale, for example, gender coding like “male = 0” and “female = 1”,
- (ii) ordinal scale, for example, ranks, its difference to a nominal scale is that the numeric coding implies, respectively, reflects, an (intentional) ordering (\leq),
- (iii) interval scale, an ordinal scale with well-defined differences, for example, temperature in $^{\circ}\text{C}$,
- (iv) ratio scale, an interval scale with true zero point, for example, temperature in $^{\circ}\text{K}$,
- (v) absolute scale, a ratio scale with (absolute) prefixed unit size, for example, inhabitants.

Let us first look at the difference between a ratio and an interval scale: the true or absolute zero point enables statements like “ 20°K is twice as warm/hot than 10°K ” to make sense while the same statement for 20°C and 10°C holds relative to the $^{\circ}\text{C}$ -scale only but not “absolute” since $293,15^{\circ}\text{K}$ is not twice as “hot” as $283,15^{\circ}\text{K}$. Interval scales allow valid statements like: let temperature on day A = 25°C , on day B = 15°C , and on day C = 20°C . Now the ratio $(A-B)/(A-C) = 2$ validates “The temperature difference between day A and B is twice as much as between day A and day C”.

As mentioned in the previous sections, nominal scale clustering allows nonparametric methods or already (distribution free) principal component analysis likewise approaches. Examples of nominal and ordinal scaling are provided in [29]. A distinction of ordinal scales into ranks and scores is outlined in [30]. While ranks just provide an ordering relative to the other items under consideration only, scores are enabling a more precise idea of distance and can have an independent meaning. In case that a score in fact has an independent meaning, that is, meaningful usability not only in case of the items observed but by an independently defined difference, then a score provides an interval scale. An ordering is called strict if and only if “ $<$ ” holds.

Example 2 (Rank to score to interval scale). Let us evaluate the response behavior of an IT-system. The evaluation answers ranked according to a qualitative ordinal judgement scale are

deficient (failed) $<$ acceptable (partial) $<$ comfortable (compliant).

Now let us assign “acceptance points” p to construct a score of “weighted ranking”:

deficient = $2p$ $<$ acceptable = $7p$ $<$ comfortable = $10p$.

This gives an idea of (subjective) distance: 5 points needed to reach “acceptable” from “deficient” and further 3 points to reach “comfortable”. But from an interpretational point of view, an interval scale should fulfill that the five points from “deficient” to “acceptable” are in fact 5/3 of the three points from “acceptable” to “comfortable” (well-defined)

and that the same score is applicable at other IT-systems too (independency). Therefore consider, as “throughput” measure, time savings:

“deficient” = loosing more than one minute = -1 ,
 “acceptable” = between loosing one minute and gaining one = 0 ,
 “comfortable” = gaining more than one minute = 1 .

For a fully well-defined situation, assume context constrains so that not more than two minutes can be gained or lost. So from “deficient” to “comfortable”, the distance will always be “two minutes”.

3.1. Transforming Ordinal Scales into Interval Scales

Lemma 1. Each strict score $(s_i)_I$ with finite index set I can be bijectively transformed into an order preserving ranking $(r_i)_I$ with $r_i = i$.

Proof. Since the index set is finite $I = \{1, \dots, n\}$ is a valid representation of the index set and the strict ordering provides s_1 to be the minimal scoring value with $s_i < s_j$ if and only if $i < j$. Thus $s_i \rightarrow i$ is the desired mapping. \square

Aside of the rather abstract “ $s_i \rightarrow i$ ”, there is a calculus of the weighted ranking $(r_i)_I = (w_i \cdot t_i)_I$ with $w_i = i/t_i$ and $t_i = s_i + |s_1| + 1$ which is order preserving and since $t_i \geq 1$ for all i it provides the desired (natural) ranking $r_i = i$. Of course qualitative expressions might permit two or more items to occupy equal rank in an ordered listing but with assigning numeric values differentiation aspects are lost if different items represented by the same numeral. Approaches to transform (survey) responses expressed by (non metric) judges on an ordinal scale to an interval (or synonymously “continuous”) scale to enable statistical methods to perform quantitative multivariate analysis are presented in [31]. Thereby a transformation-based on the decomposition into orthogonal polynomials (derived from certain matrix products) is introduced which is applicable if equally spaced integer valued scores, so-called natural scores, are used. Also the principal transformation approaches proposed from psychophysical theory with the original intensity C as judge evaluation are mentioned there.

Fechner’s law. $R = l \cdot \log(C)$ with constant l in $]0, 1[$.

Steven’s Power Law. $R = l \cdot C^n$ where l depends on the number of units and n is a measure of the rate of growth of perceived intensity as a function of stimulus intensity.

The Beidler Model. $R = C/(l - C)$ with constant l usually close to 1.

Thereby the determination of the constants or that the original ordering is lost occurs to be problematic. From lemma1 on the other-hand we see that given a strict ranking of ordinal values only, additional (qualitative context) constrains might need to be considered when assigning

a numeric representation. Of course each such condition will introduce tendencies. So without further calibration requirements it follows:

Consequence 1. An equidistant interval scaling $S = (s_i)$ which is symmetric and centralized with respect to expected scale mean $E(S) = \mu_S$ is minimizing dispersion and skewness effects of the scale.

Proof. If $\mu_S \neq 0$, let $S' = (s_i - \mu_S)$. Since

$$E(S) = \mu_S = \frac{1}{r} \sum_{i=1}^r s_i = 0 \quad (7)$$

and the symmetry condition holds for each s_i , there exist an s_j with $s_j = -s_i$. Thus the centralized second momentum reduces to

$$\sum (s_i - \mu_S)^2 = \sum s_i^2 \quad (8)$$

and the third, since $s_i^3 = -s_j^3$, to

$$\sum (s_i - \mu_S)^3 = \sum s_i^3 = 0. \quad (9)$$

□

Remark 1. For $\mu_S \neq 0$, the symmetry condition (for s_i there is an s_j with $s_i = -s_j$) reduces the centralized second momentum to

$$\sum (s_i - \mu_S)^2 = \sum (s_i^2 - 2s_i \cdot \mu_S + \mu_S^2) = \sum (s_i^2 + \mu_S^2). \quad (10)$$

In case of a strict score even to

$$\sum_{i=1}^{|I|} (s_i^2 + \mu_S^2) = 2 \cdot \sum_{i=1}^{\lfloor |I|/2 \rfloor} (s_i^2 + \mu_S^2). \quad (11)$$

Example 3. Let us return to the samples of Example 1. The predefined answer options are “fully compliant (C)”, “partial compliant (P)”, “failed (F)”, and “not applicable (N/A)”. In fact it turns out that the participants add a fifth namely, “no answer” = “blank”. A qualitative view gives $C > P > N/A > F$ since N/A should be neither positive nor negative in impact whereas F indicates a high probability of negative impact. The interpretation of “no answer” tends to be rather nearby F than at N/A—“not considered” is rather “failed” than a sound N/A judgment. Finally to assume $F >$ “blank” or $F <$ “blank” is a qualitative (context) decision. So due to the odd number of values the scaling, $C = 1$, $P = 0,5$, $N/A = 0$, “blank” = $-0,5$, and $F = -1$ may hold. In sense of a qualitative interpretation, a 0-1 (nominal) only answer option does not support the valuation mean (0,5) as an answer option and might be considered as a class predifferentiator rather than as a reliable detail analysis base input.

Corollary 1. Each (strict) ranking $\{r_i\}_I = \{r_{\min}, \dots, r_{\max}\}$, and so each score, can be consistently mapped into $[0, 1]$ via $r_i \rightarrow (r_i - r_{\min}) / (r_{\max} - r_{\min})$.

Proof. Clearly

$$f(x) = \frac{x - r_{\min}}{r_{\max} - r_{\min}} = \frac{1}{r_{\max} - r_{\min}} \cdot x - \frac{r_{\min}}{r_{\max} - r_{\min}} \quad (12)$$

is strictly monotone increasing since $f'(x) = 1 / (r_{\max} - r_{\min}) > 0$ and it gives $f(r_{\min}) = 0$ and $f(r_{\max}) = 1$. Furthermore, $f(E(x)) = E(f(x))$ and $\text{Var}(ax + b) = a^2 \text{Var}(x)$ for the variance under linear $f(x)$ shows the consistent mapping of σ -ranges. □

Remark 2. Generally such target mapping interval transformations can be viewed as a “microscope effect” especially if the inverse f^{-1} mapping from $[0, 1]$ into a larger interval is considered. Qualitative interpretations of the occurring values have to be done carefully since it is not a representation on a ratio or absolute scale. Similar magnifying effects are achievable by applying power or root functions to values out of interval $[0, 1]$.

Remark 3. The values v_k out of $[0, 1]$ associated to (ordinal) rank r_k are not the probabilities of occurrence. Let us look again at Examples 1 and 3. Each sample event x_{ij} is mapped onto a value v_k ($k = 1, \dots, r$; here $r = 5$). Thus each $x_{ij} = v_k$ with k depending on (i, j) . Let $a_k = |\{(i, j) \mid x_{ij} = v_k\}|$ denote the total number of occurrence of v_k and let the full sample X with $|X| = n \cdot m$. Then the (empirical) probability $P(\cdot)$ of occurrence of v_k is expressed by $p_k = P(x_{ij} = v_k) = a_k / n \cdot m$. In fact

$$\sum_{k=1}^r p_k = \sum_{k=1}^r \frac{a_k}{n \cdot m} = \frac{1}{n \cdot m} \sum_{k=1}^r a_k = \frac{|X|}{n \cdot m} = 1,$$

$$\mu = E(X) = \sum_{k=1}^r v_k \cdot p_k = \frac{1}{n \cdot m} \sum_{k=1}^r v_k a_k = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m x_{ij}. \quad (13)$$

Analog with $a_k^{(i)}$ as the total of v_k occurrence at the sample block of question i ,

$$\mu_i = \sum_{k=1}^r v_k p_k^{(i)} = \sum_{k=1}^r v_k \frac{a_k^{(i)}}{m}. \quad (14)$$

And thus it gives p_k as the expected mean of $p_k^{(i)}$

$$\mu = \frac{1}{n} \sum_{i=1}^n \mu_i = \sum_{k=1}^r \left(v_k \cdot \frac{1}{n} \sum_{i=1}^n p_k^{(i)} \right) \Rightarrow \forall k : p_k = \frac{1}{n} \sum_{i=1}^n p_k^{(i)}. \quad (15)$$

Remark 4. The essential empiric mean equation “ $\mu = \sum v_k p_k$ ” is nicely outlining the intended weighting through the actual occurrence a_k of the value v_k but also that even a weak symmetry condition only, like $\sum_{v_k \leq 0} |v_k| = \sum_{v_k \geq 0} v_k$, might already cause an inappropriate bias.

4. Analysis Modelling

The key to analysis approaches in spite of determining areas of potential improvements is an appropriate underlying

model providing reasonable theoretical results which are compared and put into relation to the measured empirical input data.

4.1. Remarks on the Reliability of Statistical Result Analysis. Perhaps the most frequent assumptions mentioned when applying mathematical statistics to data are the Normal distribution (Gauß' bell curve) assumption and the (stochastic) independency assumption of the data sample (for elementary statistics see, e.g., [32]). For both a χ^2 -test can be utilized. The Normal-distribution assumption is also coupled with the sample size. If the sample size is huge enough the central limit theorem allows assuming Normal-distribution or at smaller sizes a Kolmogoroff-Smirnoff test may apply or an appropriate variation. The Normal-distribution assumption is utilized as a base for applicability of most of the statistical hypothesis tests to gain reliable statements. The independency assumption is typically utilized to ensure that the calculated estimation values are usable to reflect the underlying situation in an unbiased way. The χ^2 -independency testing is realized with contingency tables. In any case it is essential to be aware about the relevant testing objective.

What are we looking for being normally distributed in Example 1 and why? A quite direct answer is "looking for the distribution of the answer values to be used in statistical analysis methods". Thereby the marginal mean values of the questions

$$\bar{\mu}_Q = (\mu_1, \dots, \mu_n)^T \quad \text{with } \mu_i = \mu_{i\bullet} = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad (16)$$

as well as the marginal mean values of the surveys in the sample

$$\bar{\mu}_S = (\mu_{S_1}, \dots, \mu_{S_m})^T \quad \text{with } \mu_{S_j} = \mu_{\bullet j} = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (17)$$

are showing up $\mu = (1/(n \cdot m)) \sum_{i=1}^n \sum_{j=1}^m x_{ij}$ as the overall mean value (cf. (2)).

Applying a Kolmogoroff-Smirnoff test at the marginal means forces the selected scoring values to pass a validity check with the tests allocated α -significance level. That is, if the Normal-distribution hypothesis cannot be supported on significance level α , the chosen valuation might be interpreted as inappropriate. Briefly the maximum difference of the marginal means cumulated ranking weight (at descending ordering the [total number of ranks minus actual rank] divided by total number of ranks) and their expected $N(\mu_i | \mu, \sigma_Q)$ result should be small enough, for example, for $\alpha = 0,05$ lower than $1,36/\sqrt{n}$ and for $\alpha = 0,01$ lower than $1,63/\sqrt{n}$. For $n = 104$ this evolves to (rounded) 0,13, respectively, 0,16 ($\alpha = 0,01$). In case of Example 3 and initial reviews the maximum difference appears to be 0,095. Thus for $\alpha = 0,01$ the Normal-distribution hypothesis is acceptable. In case of switching F and "blank", it shows 0,09 as calculated maximum difference. In case of C = 1, P = 0,7, F = 0, and N/A and "blank" not counted, the maximum difference is 0,29 and so the Normal-distribution hypothesis

has to be rejected for $\alpha = 0,05$ and $\alpha = 0,01$, that is, neither an inappropriate rejection of 5% nor of 1% of normally distributed sample cases allows the general assumption of Normal-distribution hypothesis in this case. The same test results show up for the case study with the μ_S -type marginal means ($m = 37$). The symmetry of the Normal-distribution and that the interval $[\mu - \sigma, \mu + \sigma]$ contains $\sim 68\%$ of observed values are allowing a special kind of "quick check": "if $\mu \pm \sigma$ exceeds the sample values at all, the Normal-distribution hypothesis should be rejected." which appears in the case study at the N/A and "blank" not counted case. The statistical independency of random variables ensures that calculated characteristic parameters (e.g., unbiased estimators) allow a significant and valid interpretation. To apply χ^2 -independency testing with $(|K| - 1) \cdot (|L| - 1)$ degrees of freedom, a contingency table $H = (h_{kl})_{K \times L}$ with h_{kl} counting the common occurrence of observed characteristic k out of index set K and l out of index set L is utilized and as test statistic (h_{\bullet} indicates a marginal sum; $h_{\bullet\bullet} = h$)

$$\chi_T^2 = \sum_K \sum_L \frac{(h_{kl} - h_{kl}^*)^2}{h_{kl}^*} \quad \text{with } h_{kl}^* = \frac{h_{k\bullet} \cdot h_{\bullet l}}{h}, \quad (18)$$

$$h = \sum_K h_{k\bullet} = \sum_L h_{\bullet l}.$$

So on significance level α the independency assumption has to be rejected if $\chi_T^2 > \chi^2(1 - \alpha; (m - 1)(r - 1))$ for the $(1 - \alpha)$ quantile of the χ^2 -distribution.

Looking at the case study the colloquial "the answers to the questionnaire should be given independently" needs to be stated more precisely. Of course independency can be checked for the gathered data project by project as well as for the answers by appropriate χ^2 -tests. In case of the project by project level the independency of project A and project B responses can be checked with h_{kl} as the count of "answers with value "k" at project A and answer value "l" at project B". Thus is that independency telling us that one project is not giving an answer because another project has given a specific answer. So a distinction and separation of timeline given repeated data gathering from within the same project is recommendable.

In case of the answers in-between relationship, it is neither a priori intended nor expected to have the questions and their results always statistically independent, especially not if they are related to the same superior procedural process grouping or aggregation. It is even more of interest how strong and deep a relationship or dependency might be. In case of normally distributed random variables it is a well-known fact that independency is equivalent to being uncorrelated (e.g., [32]). Thereby, the (Pearson-) correlation coefficient of X and Y is defined through $\rho_{XY} = \sigma_{XY}/\sigma_X\sigma_Y$ with σ_X , σ_Y as the standard deviation of $X = (x_i)$, respectively, $Y = (y_i)$ and σ_{XY} as their covariance

$$\sigma_{XY} = \text{Cov}(X, Y) = \frac{1}{n - 1} \sum_{i \in I} (x_i - \mu_X)(y_i - \mu_Y). \quad (19)$$

Thus it allows also a "quick check/litmus test" for independency: "if the (empirical) correlation coefficient exceeds

a certain value the independency hypothesis should be rejected". Of course there are also exact tests available for ρ , for example, for $H_0: \rho = 0$ from a t -distribution test statistic or from the normal distribution $N((1/2) \ln((1 + \rho_0)/(1 - \rho_0)); 1/(n - 3))$ with ρ_0 as the real value [32].

4.2. *Extended Modelling with Correlation Coefficients.* In sense of our case study, the straight forward interpretation of the answer correlation coefficients—note that we are not taking the Spearman's rho here—allows us to identify questions within the survey being potentially obsolete ($\rho_{xy} = 1$) or contrary ($\rho_{xy} = -1$). The expressed measure of linear dependency is pointing out overlapping areas ($\rho_{xy} > 0$) or potential conflicts ($\rho_{xy} < 0$). Aside of this straight forward usage, correlation coefficients are also a subject of contemporary research especially at principal component analysis (PCA); for example, as earlier mentioned in [23] or at the analysis of hebbian artificial neural network architectures whereby the correlation matrix' eigenvectors associated with a given stochastic vector are of special interest [33]. In conjunction with the α -significance level of the coefficients testing, some additional meta-modelling variables may apply. So the absolute value of recognized correlation coefficients may have to exceed a defined lower limit before taken into account; aggregation within specified value ranges of the coefficients may be represented by the ranges mean values; the signing as such may be ignored or combinations of these options are possible.

At least in situations with a predefined questionnaire, like in the case study, the single questions are intentionally assigned to a higher level of aggregation concept, that is, not only PCA will provide grouping aspects but there is also a predefined intentional relationship definition existing. In our case study, these are the procedures of the process framework. Such (qualitative) predefined relationships are typically showing up the following two quantifiable construction parameters:

- (i) a weighting function outlining the "relevance" or "weight" of the lower level object, relative within the higher level aggregate,
- (ii) the number of allowed low to high level allocations.

For example, such an initial relationship indicator matrix $I_{p,n}$ for procedures ($p < n$) given per row and the allocated questions as columns with constant weight $w = 1$, interpreted as fully adhered to the indicated allocation, and with a (directed) 1:1 question-procedure relation, as a "primary main procedure allocation" for the questions, will give, if ordered appropriate, a somewhat diagonal block relation structure:

$$I = \begin{pmatrix} 1 \dots 1 & 0 & \dots & \dots & 0 \\ 0 & 1 \dots 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \dots 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \dots 1 \end{pmatrix}_{p,n}. \quad (20)$$

An approach to receive value from both views is a model combining the (experts) presumable indicated weighted relation matrix $I_{p,n}$ with the empirically determined PCA relevant correlation coefficients matrix $\text{Cor}(X, X) = (\rho_{ij})_{n,n}$.

Such a scheme is described by the linear aggregation modelling of the form

$$A_{p,n} = I_{p,n} \cdot \text{Cor}(X, X) = (a_{ij})_{p,n}. \quad (21)$$

Additional to the meta-modelling variables "magnitude and validity of correlation coefficients" and applying "value range means representation" to the matrix multiplication result, a normalization transformation $T : \mathfrak{R} \rightarrow [-1, 1]$ appears to be expedient. Some obvious but relative normalization transformations are disputable:

(1)

$$T^{(1)} : a_{ij} \mapsto t_{ij} = \frac{a_{ij}}{\max\{|a_{is}| : s = 1, \dots, n\}} \quad (22)$$

for each i fixed,

This might be interpreted that the $|a_{ij}| = \max\{|a_{is}| : s = 1, \dots, n\}$ will be 100% relevant to aggregate in row i but there is no reason to assume in case of $|a_{kj}| > |a_{ij}|$ that the column object j being less than 100% relevant to aggregate k which happens if the maximum in row k is greater than $|a_{kj}|$.

(2) Also the

$$T^{(2)} : a_{ij} \mapsto t_{ij} = \frac{a_{ij}}{\sum_j a_{ij}} \quad (23)$$

transformation is indeed keeping the relative portion within the aggregates and $\sum_j t_{ij} = \sum_j (a_{ij} / \sum_j a_{ij}) = (1 / \sum_j a_{ij}) \cdot \sum_j a_{ij} = 1$ might be interpreted as 100% coverage of the row aggregate through the column objects but it assumes collaterally disjunct coverage by the column objects too.

(3) Most appropriate in usage and similar to eigenvector representation in PCA is the normalization via the (Euclidean) length $l(x) = \sqrt{\langle x, x \rangle} = \sqrt{x \cdot x} = \sqrt{\sum_{s \in S} x_s^2}$, that is, in relation to the aggregation object i and the row vector $r_i = (a_{i1}, \dots, a_{in})$, the transformation

$$T^{(3)} : a_{ij} \mapsto t_{ij} = \frac{a_{ij}}{l(r_i)} \quad (24)$$

yields, since the length of the resulting row vector equals 1, a 100% interpretation coverage of aggregate i , providing the relative portions and allowing conjunctive input of the column defining objects.

Now with E as the unit-matrix and $X \sim N(\bar{\mu}, \text{Cov}(X, X))$, we can assume

$$\beta = \beta(A)_{p,1} := I_{p,n} \cdot \text{Cor}(X, X)_{n,n} \cdot \bar{\mu}_{n,1} + \bar{\varepsilon} \quad (25)$$

with standard error $\vec{\varepsilon} \sim N(0, E)$ as the aggregation level built-up statistical distribution model $M_{p,n}$ (e.g., questions \rightarrow procedures). In terms of the case study, the aggregation to procedure level built-up model-based on given answer results is expressible as (see (24) and (25))

$$\beta(T^{(3)}(A))_{p,1} = T^{(3)}(I_{p,n} \cdot \text{Cor}(Q, Q)_{n,n}) \cdot \vec{\mu}_Q + \vec{\varepsilon}. \quad (26)$$

Notice that with transformation $T^{(3)}$ applied and since $|\mu_j| \leq 1$ implies $|t_{ij} \cdot \mu_j| \leq |t_{ij}|$ it holds

$$\begin{aligned} \sum_j (t_{ij}\mu_j)^2 &\leq \sum_j |t_{ij}| |t_{ij} \cdot \mu_j| \\ &\leq \sum_j |t_{ij}| |t_{ij}| = \sum_j (t_{ij})^2 = 1. \end{aligned} \quad (27)$$

Recall $|\mu_j| \leq 1$ will be a natural result if the underlying scaling is from within $[-1, 1]$. If appropriate, for example, for reporting reason, β might be transformed according $l(\beta) = 1$ or according to Corollary 1. Also notice that $I = E$ matches with the common PCA modelling base. With Ψ as an eigenvector associated with eigen-value λ of $\text{Cor}(X, X)$ an idealized heuristic ansatz to measure consilience results in

$$\begin{aligned} (I \cdot \text{Cor}(X, X) - \lambda E) \cdot \Psi &= I \cdot ((\rho_{XY}) \cdot \Psi) - \lambda \Psi \\ &= (I - E) \cdot \lambda \cdot \Psi. \end{aligned} \quad (28)$$

So under these terms the “difference” of the model compared to a PCA model is depending on $(I - E)$. This points into the direction that a predefined indicator matrix aggregation equivalent to a more strict diagonal block structure scheme might compare better to a PCA empirically derived grouping model than otherwise (cf. also topological ultra-filters in [15]). This is comprehensible because of the orthogonality of the eigenvectors but there is not necessarily a component-by-component disjunction required. In fact a straight forward interpretation of the correlations might be useful but for practical purpose and from practitioners view a referencing of only maximal aggregation level is not always desirable. So for evaluation purpose ultrafilters, multilevel PCA sequence aggregations (e.g., in terms of the case study: PCA on questions to determine procedures—PCA on procedures to determine processes—PCA on processes to determine domains, etc.) or too broadly-based predefined aggregation might avoid the desired granularity for analysis.

4.3. Adherence of Gathered Data to Aggregation Model. In order to answer how well observed data will adhere to the specified aggregation model it is feasible to calculate the aberration as a function induced by the empirical data and the theoretical prediction. Formally expressed through

$$\text{Adh}(M_{p,n}, X) = F(A_{p,n}, \vec{\mu}_X). \quad (29)$$

Thereby the adherence(Adh) to a single aggregation form F_i (i in $\{1, \dots, p\}$) is of interest. So let $\mathbf{F} = (F_1, \dots, F_p)^T$ whereby $F_i : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}$ is the calculation result of a comparison of the aggregation represented by the i th row-vector \mathbf{r}_i of $A_{p,n}$ ($= r_i(A)$) and the effect triggered by the observed $\mu_X = (\mu_1, \dots, \mu_n)^T$. Notice that in the notion of the case study μ_Q is considered and $\mu_1 := (1, \dots, 1)^T$ equals “everything is fully compliant” with no aberration and $|\mu_j| \leq 1$ holds. So options of \mathbf{F} are given through

- (1) $\beta_i = \sum_j a_{ij} \cdot \mu_j (= r_i(A) \cdot \mu_X)$ compared to $\sum_j a_{ij} (= r_i(A) \cdot \mu_1)$ and adherence formula:

$$F_i^{(1)} : (r_i(A), \vec{\mu}_X) \mapsto \frac{\beta_i}{\sum_j a_{ij}} = \frac{\langle r_i(A), \vec{\mu}_X \rangle}{\langle r_i(A), \vec{\mu}_1 \rangle} \quad (30)$$

but this can be formally only valid if β_i and $\sum_j a_{ij}$ have the same sign since the theoretical $\min(\text{Adh}) = 0$ expresses already fully incompliance.

- (2) Let $*$ denote a component-by-component multiplication so that $A_{p,n} * \mu_X = B_{p,n} = ([b_{ij} = a_{ij} \cdot \mu_j])_{p,n}$. Similarly as in (30) an adherence measure-based on disparity (in sense of a “length” compare) is provided by

$$\begin{aligned} F_i^{(2)} : (r_i(A), \vec{\mu}_X) &\mapsto \frac{l(r_i(A * \vec{\mu}_X))}{l(r_i(A))} \\ &= \sqrt{\frac{\langle r_i(A * \vec{\mu}_X), r_i(A * \vec{\mu}_X) \rangle}{\langle r_i(A), r_i(A) \rangle}} \\ &= \sqrt{\frac{\sum_j (a_{ij} \cdot \mu_j)^2}{\sum_j a_{ij}^2}}. \end{aligned} \quad (31)$$

Notice that $\mu_X = \mu_1 = (1, \dots, 1)^T$ gives $A_{p,n} = B_{p,n}$.

- (3) An azimuth measure of the angle θ between $r_i(A * \mu_X)$ and $r_i(A)$

$$\begin{aligned} F_i^{(3)} : (r_i(A), \vec{\mu}_X) &\mapsto \langle r_i(A), r_i(A * \vec{\mu}_X) \rangle \\ &= \sum_j a_{ij}^2 \mu_j = l(r_i(A)) \cdot l(r_i(A * \vec{\mu}_X)) \cdot \cos(\theta). \end{aligned} \quad (32)$$

The orientation of the vectors in the underlying vector space, that is, simply spoken “if a vector is on the “left” or “right” side of the other”, does not matter in sense of adherence measurement and is finally evaluated by an examination analysis of the single components characteristics. In addition the constrain $\max(\text{Adh}) = 1$, that is, full adherence, has to be considered too. So let us specify \mathbf{F} under assumption

$T = T^{(3)}(A)$ and with $|\mu_j| \leq 1$ as a consequence from scaling values out of $[-1, 1]$:

$$\beta = (T * \vec{\mu}_X) \cdot \vec{\mu}_1, \quad \beta_i = \sum_j t_{ij} \mu_j, \quad (33)$$

$$|\beta_i| \leq \sum_j |t_{ij} \mu_j| \leq \sum_j |t_{ij}|,$$

$$F_i^{(1)}(r_i(T), \vec{\mu}_X) \in [0, 1] \iff |\beta_i| \leq \left| \sum_j t_{ij} \right| \wedge \text{sgn}(\beta_i) = \text{sgn}\left(\sum_j t_{ij}\right), \quad (34)$$

$$F_i^{(2)} = F_i^{(2)}(r_i(T), \vec{\mu}_X) = \sqrt{\sum_j (t_{ij} \mu_j)^2} \leq 1 \quad \text{cf. (27)}. \quad (35)$$

Recall that the following generally holds

$$\sigma_X^2 = E(X^2) - E(X)^2. \quad (36)$$

Thus for $Y = \{r_i(T) * \vec{\mu}_X\}_J = \{t_{ij} \cdot \mu_j\}_{j=1, \dots, n}$ we get

$$\sigma_Y^2 = \frac{1}{n} \sum_j (t_{ij} \mu_j)^2 - \left(\frac{1}{n} \sum_j (t_{ij} \mu_j) \right)^2 \implies (F_i^{(2)})^2 = n \cdot \sigma_Y^2 + \frac{1}{n} \beta_i^2. \quad (37)$$

Since θ and $\cos(\theta)$ are independent from the length of the examined vectors, we might apply $T = T^{(3)}(A)$ and $T^{(3)}(A * \mu_X)$. So let $l(r_i(T)) = l(r_i(T^{(3)}(A * \mu_X))) = 1$. And since $T^{(3)}(A * \mu_X) = T^{(3)}(T^{(3)}(A) * \mu_X)$ holds, which is shown by

$$\begin{aligned} & T^{(3)}(r_i(T^{(3)}(A) * \vec{\mu}_X)) \\ &= T^{(3)}(r_i(T) * \vec{\mu}_X) \\ &= \frac{1}{l(r_i(T * \vec{\mu}_X))} (\dots, t_{ij} \mu_j, \dots) \\ &= \frac{(1/l(r_i(A))) (\dots, a_{ij} \mu_j, \dots)}{l(r_i(A * \vec{\mu}_X)/l(r_i(A)))} \\ &= \frac{(1/l(r_i(A))) (\dots, a_{ij} \mu_j, \dots)}{(1/l(r_i(A))) l(r_i(A * \vec{\mu}_X))} \\ &= \frac{(\dots, a_{ij} \mu_j, \dots)}{l(r_i(A * \vec{\mu}_X))} = T^{(3)}(r_i(A * \vec{\mu}_X)). \end{aligned} \quad (38)$$

thus $F^{(3)}$ evolves to

$$\begin{aligned} F_i^{(3)}(r_i(T), \vec{\mu}_X) &= \left| \langle r_i(T), r_i(T^{(3)}(A * \vec{\mu}_X)) \rangle \right| \\ &= \frac{1}{l(r_i(T * \vec{\mu}_X))} \left| \sum_j t_{ij}^2 \mu_j \right| \\ &= |\cos(\theta)| \in [0, 1]. \end{aligned} \quad (39)$$

Simultaneous appliance of $F^{(2)}$ and $F^{(3)}$ will give a kind of cross “check & balance” to validate and complement each other as adherence metric and measurement.

The relevant areas to identify high and low adherence results are defined by not being inside the interval (mean \pm standard deviation). An interpretation as an expression of percentage or prespecified fulfillment goals are doubtful for all metrics without further calibration specification other than 100% equals fully adherent and 0% is totally in-compliant (cf., Remark 2). The same high-low classification of value-ranges might apply to the set of the β_i . But the interpretation of a β_i is more to express the observed weight of an aggregate within the full set of aggregates than to be a compliance measure of fulfilling an explicit aggregation definition.

4.4. Model Evaluation. In contrast to the model inherit characteristic adherence measure, the aim of model evaluation is to provide a valuation base from an outside perspective onto the chosen modelling. In [34] Müller and Supatgiat described an iterative optimisation approach to evaluate compliance and/or compliance inspection cost applied to an already given effectiveness-model (indicator matrix) of “measures/influencing factors” determining “(legal regulatory) requirements/classes” as aggregates. It was also mentioned by the authors there that it took some hours of computing time to calculate a result. In fact the situation to determine an “optimised” aggregation model is even more complex. Let us recall the defining modelling parameters:

- (i) the definition of the applied scale and the associated scaling values,
- (ii) relevance variables of the correlation coefficients ($|\rho| \geq \text{constant} \ \& \ \alpha\text{-level}$),
- (iii) the definition of the relationship indicator matrix I ,
- (iv) entry value range adjustments applied to $(a_{ij}) = A = I \cdot \text{Cor}$.

Instead of a straight forward calculation, a measure of congruence alignment suggests a possible solution. Therefore, the observation result vectors μ_X and β will be compared with the modeling inherit expected theoretical estimated values derived from the model matrix A . Under the assumption that the modeling is reflecting the observed situation sufficiently the appropriate localization and variability parameters should be congruent in some way. That is, the appliance of a well-defined value transformation will provide the possibility for statistical tests to decide if the observed and the theoretic outcomes can be viewed as samples from within the same population.

Let

$$\vec{\mu}_X \in [-1, 1]^n, \quad \beta = (T * \vec{\mu}_X) \cdot \vec{\mu}_1 \in \mathfrak{R}^p \wedge \vec{\mu}_1 = (1)_{n,1} \quad (40)$$

be the observed values and

$$\tilde{\mu}_X = \frac{1}{l(\vec{\mu}_X)} \vec{\mu}_X = (\tilde{\mu}_i) \wedge \tilde{\beta} = \frac{1}{l(\beta)} \beta = (\tilde{\beta}_j) \quad (41)$$

representing the uniquely transformed values.

Analog the theoretic model estimating values are expressed as ($T^T = "T$ transposed")

$$\begin{aligned} \hat{\mu}_X &= \frac{1}{l(T^T \cdot \hat{\mu}_1)} T^T \cdot \hat{\mu}_1 = (\hat{\mu}_i)_{n,1} \wedge \hat{\mu}_1 = (1)_{p,1}, \\ \hat{\beta} &= \frac{1}{l(T \cdot \vec{\mu}_1)} T \cdot \vec{\mu}_1 = (\hat{\beta}_j)_{p,1}. \end{aligned} \quad (42)$$

Now the relevant statistical parameter values are

$$\begin{aligned} \tilde{\mu} &= \frac{1}{n} \sum_I \tilde{\mu}_i \wedge \tilde{\sigma}^2 = \frac{1}{n-1} \sum_I (\tilde{\mu}_i - \tilde{\mu})^2, \\ \hat{\mu} &= \frac{1}{n} \sum_I \hat{\mu}_i \wedge \hat{\sigma}^2 = \frac{1}{n-1} \sum_I (\hat{\mu}_i - \hat{\mu})^2, \\ \tilde{\beta} &= \frac{1}{p} \sum_J \tilde{\beta}_j \wedge \tilde{\sigma}_\beta^2 = \frac{1}{p-1} \sum_J (\tilde{\beta}_j - \tilde{\beta})^2, \\ \hat{\beta} &= \frac{1}{p} \sum_J \hat{\beta}_j \wedge \hat{\sigma}_\beta^2 = \frac{1}{p-1} \sum_J (\hat{\beta}_j - \hat{\beta})^2. \end{aligned} \quad (43)$$

The evaluation is now carried out by performing statistical significance testing for

$$\begin{aligned} [H_0 : \tilde{\mu} = \hat{\mu}] \wedge [H'_0 : \tilde{\sigma}^2 = \hat{\sigma}^2], \\ [H_0^\beta : \tilde{\beta} = \hat{\beta}] \wedge [H'_0{}^\beta : \tilde{\sigma}_\beta^2 = \hat{\sigma}_\beta^2] \end{aligned} \quad (44)$$

with the corresponding H_1 hypothesis. Of course thereby the probability $(1-\alpha)$ under which the hypothesis is valid is of interest. So not a test result to a given significance level α is to be calculated but the minimal α (or percentile) under which the H_0 hypothesis still holds. The appropriate test statistics on the means (μ, β) are according to a (two-tailed) Student's t -distribution and on the variances (σ^2) according to a Fisher's F -distribution. For practical purpose the desired probabilities are ascertainable, for example, with spreadsheet program built-in functions "TTEST" and "FTEST" (e.g., Microsoft Excel, IBM Lotus Symphony, SUN Open Office). Reasonable varying of the defining modelling parameters will therefore provide t -test and F -test results for the direct observation data (H_0, H'_0) and for the aggregation objects $(H_0^\beta, H'_0{}^\beta)$. The ultimate goal is that all probabilities are tending towards 1. But this is quite unrealistic and a decision of accepting a model set-up has to take surrounding qualitative perspectives too. So it

might occur that an improved concordance at the aggregates is coupled with a decrease of a probability value at the observation data side or any other uncomfortable situation depending on which of the defining variables is changed. Especially the aspect to use the model theoretic results as a base for improvement recommendations regarding aggregate adherence requires a well-balanced adjustment and an overall rating at a satisfactory level. As a rule of thumb a well-fitting localizing t -test value at the observed data is considerable more valuable than the associated F -test value since a correct predicted mean looks more important to reflect coincidence of the model with reality than a prediction of the spread of individual triggered responses. A little bit different is the situation for the aggregates level. Since the aggregates are artificially to a certain degree the focus of the model may be at explaining the variance rather than at the average localization determination but with a tendency for both values at a similar magnitude. As an illustration of input/outcome variety the following changing variables value sets applied to the case study data may be considered to shape on a potential decision issue (t - and F -test values with $Q =$ Question, $P =$ aggregating procedure):

- (i) a (specified) matrix \mathbf{I} with entries either 0 or 1; $|\rho| \geq 0,625$ is resulting in:
 $t_Q = 0,837, \quad F_Q = 0,001, \quad t_P = 0,641, \quad F_P = 0,659.$ (45)
- (ii) \mathbf{I} as above but with entries "1" substituted from $\{0.3, 0.5, 0.7, 1\}$; $|\rho| \geq 0,65$ and the entries of $\mathbf{I} \cdot \mathbf{A}$ consolidated at margin and range means $\{0; 0,35; 0,7; 1\}$:
 $t_Q = 0,77, \quad F_Q = 0,023, \quad t_P = 0,557, \quad F_P = 0,905.$ (46)

5. Conclusions and Future Research

The need to evaluate available information and data is increasing permanently in modern times. Thereby more and more qualitative data resources like survey responses are utilized. Therefore a methodic approach is needed which consistently transforms qualitative contents into a quantitative form and enables the appliance of formal mathematical and statistical methodology to gain reliable interpretations and insights which can be used for sound decisions and which is bridging qualitative and quantitative concepts combined with analysis capability. The desired avoidance of methodic processing gaps requires a continuous and careful embodiment of the influencing variables and underlying examination questions from the mapping of qualitative statements onto numbers to the point of establishing formal aggregation models which allow quantitative-based qualitative assertions and insights.

In this paper are mathematical prerequisites depicted and statistical methodology applied to address and investigate on this issue. In particular the transformation from ordinal scaling to interval scaling is shown to be optimal if equidistant and symmetric. Alternative to principal component analysis an extended modelling to describe aggregation level

models of the observation results-based on the matrix of correlation coefficients and a predefined qualitative motivated relationship incidence matrix is introduced. On such models are adherence measurements and metrics defined and examined which are usable to describe how well the observation fulfills and supports the aggregates definitions. Finally a method combining F - and t -tests to derive a decision criteria on the fitting of the chosen aggregation model is presented. This appears to be required because the multiple modelling influencing parameters are not resulting in an analytically usable closed formula to calculate an optimal aggregation model solution. Part of these meta-model variables of the mathematical modelling are the scaling range with a rather arbitrarily zero-point, preselection limits on the correlation coefficients values and on their statistical significance relevance-level, the predefined aggregates incidence matrix and normalization constraints. If some key assumption from statistical analysis theory are fulfilled, like normal distribution and independency of the analysed data, a quantitative aggregate adherence calculation is enabled. Therefore two measurement metrics namely a dispersion (or length) measurement and a azimuth (or angle) measurement are established to express quantitatively the qualitative aggregation assessments.

In fact the quantifying method applied to data is essential for the analysis and modelling process whenever observed data has to be analyzed with quantitative methods. The presented modelling approach is relatively easy implementable especially whilst considering expert-based preaggregation compared to PCA.

An important usage area of the extended modelling and the adherence measurement is to gain insights into the performance behaviour related to the not directly evaluable aggregates or category definitions. In the case study this approach and the results have been useful in outlining tendencies and details to identify focus areas of improvement and well performing process procedures as the examined higher level categories and their extrapolation into the future.

As a continuation on the studied subject a qualitative interpretations of β , a refinement of the t - and F -test combination methodology and a deep analysis of the Eigenspace characteristics of the presented extended modelling compared to PCA results are conceivable, perhaps in adjunction with estimating questions.

Acknowledgments

The author would like to acknowledge the IBM IGA Germany EPG for the case study raw data and the IBM IGA Germany and Beta Test Side management for the given support. The author also likes to thank the reviewer(s) for pointing out some additional bibliographic sources.

References

- [1] J. Neill, "Qualitative versus Quantitative Research: Key Points in a Classic Debate," 2007, <http://wilderdom.com/research/QualitativeVersusQuantitativeResearch.html>.
- [2] D. Siegle, "Qualitative versus Quantitative," <http://www.gifted.uconn.edu/siegle/research/Qualitative/qualquan.htm>.
- [3] P. Hodgson, "Quantitative and Qualitative data—getting it straight," 2003, <http://www.blueprintusability.com/topics/articlequantqual.html>.
- [4] D. M. Mertens, *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*, Sage, London, UK, 2005.
- [5] J. Neill, "Analysis of Professional Literature Class 6: Qualitative Re-search I," 2006, <http://www.wilderdom.com/OECourses/PROFLIT/Class6Qualitative1.htm>.
- [6] J. Neill, "Analysis of Professional Literature Class 4: Quantitative Research Designs: Experimental, Quasi-Experimental, & Non-Experimental," 2003, <http://www.wilderdom.com/OECourses/PROFLIT/Class4QuantitativeResearchDesigns.htm>.
- [7] M. Q. Patton, *Qualitative Research and Evaluation Methods*, Sage, London, UK, 2002.
- [8] A. Tashakkori and C. Teddlie, *Mixed Methodology: Combining Qualitative and Quantitative Approaches*, Sage, Thousand Oaks, Calif, USA, 1998.
- [9] M. Sandelowski, "Focus on research methods: combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies," *Research in Nursing and Health*, vol. 23, no. 3, pp. 246–255, 2000.
- [10] A. Jakob, "Möglichkeiten und Grenzen der Triangulation quantitativer und qualitativer Daten am Beispiel der (Re-) Konstruktion einer Typologie erwerbsbiographischer Sicherheitskonzepte," *Forum Qualitative Sozialforschung*, vol. 2, no. 1, article 20, 2001.
- [11] P. Mayring, "Combination and integration of qualitative and quantitative analysis," *Forum Qualitative Sozialforschung*, vol. 2, no. 1, article 6, 2001.
- [12] D. L. Driscoll, A. Appiah-Yeboah, P. Salib, and D. J. Rupert, "Merging qualitative and quantitative data in mixed methods research: how to and why not," *Ecological and Environmental Anthropology*, vol. 3, no. 1, pp. 19–28, 2007.
- [13] P. Z. Wang and C. Dou, "Quantitative-qualitative transformations based on fuzzy logic," in *Applications of Fuzzy Logic Technology III*, vol. 2761 of *Proceedings of SPIE*, pp. 272–275, April 1996.
- [14] R. Gascon, "Verifying qualitative and quantitative properties with LTL over concrete domains," in *Proceedings of the 4th Workshop on Methods for Modalities (MAM '05)*, Informatik-Bericht no. 194, pp. 54–61, Humboldt Universität zu Berlin, Berlin, Germany, December 2005.
- [15] F. S. Herzberg, "Judgement aggregation functions and ultraproducts," 2008, http://www.researchgate.net/publication/23960811-Judgment_aggregation_functions_and_ultraproducts.
- [16] G. Canfora, L. Cerulo, and L. Troiano, "Transforming quantities into qualities in assessment of software systems," in *Proceedings of the 27th Annual International Computer Software and Applications Conference (COMPSAC '03)*, pp. 312–319, 2003.
- [17] M. A. Kłopotek and S. T. Wierzchon, "Qualitative versus quantitative interpretation of the mathematical theory of evidence," in *Proceedings of the 10th International Symposium on Foundations of Intelligent Systems (ISMIS '97)*, Z. W. Ras and A. Skowron, Eds., vol. 1325 of *Lecture Notes in Artificial Intelligence*, pp. 391–400, Springer, Charlotte, NC, USA, October 1997.
- [18] C. Driver and G. Urga, "Transforming qualitative survey data: performance comparisons for the UK," *Oxford Bulletin of Economics and Statistics*, vol. 66, no. 1, pp. 71–89, 2004.

- [19] L. L. Thurstone, "Attitudes can be measured," *American Journal of Sociology*, vol. 33, pp. 529–554, 1928.
- [20] W. M. Trochim, *The Research Methods Knowledge Base*, 2nd edition, 2006, <http://www.socialresearchmethods.net/kb>.
- [21] J. C. Gower, "Fisher's optimal scores and multiple correspondence analysis," 1990, *Biometrics*, 46, 947-961, http://www.datatheory.nl/pdfs/90/90_04.pdf.
- [22] F. W. Young, "Quantitative analysis of qualitative data," *Psychometrika*, vol. 46, no. 4, pp. 357–388, 1981.
- [23] K. Srnka and S. Koeszegi, "From words to numbers: how to transform qualitative data into meaningful quantitative results," *Schmalenbach Business Review*, vol. 59, pp. 29–57, 2007.
- [24] P. Rousset and J.-F. Giret, "Classifying qualitative time series with SOM: the typology of career paths in France," in *Proceedings of the 9th International Work-Conference on Artificial Neural Networks (IWANN '07)*, vol. 4507 of *Lecture Notes in Computer Science*, pp. 757–764, Springer, San Sebastián, Spain, June 2007.
- [25] S. K. M. Wong and P. Lingras, "Representation of qualitative user preference by quantitative belief functions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, no. 1, pp. 72–78, 1994.
- [26] D. Janetzko, "Processing raw data both the qualitative and quantitative way," *Forum Qualitative Sozialforschung*, vol. 2, no. 1, article 11, 2001.
- [27] D. Kuiken and D. S. Miall, "Numerically aided phenomenology: procedures for investigating categories of experience," *Forum Qualitative Sozialforschung*, vol. 2, no. 1, article 15, 2001.
- [28] H. Witt, "Forschungsstrategien bei quantitativer und qualitativer Sozialforschung," *Forum Qualitative Sozialforschung*, vol. 2, no. 1, article 8, 2001.
- [29] D. P. O'Rourke and T. W. O'Rourke, "Bridging the qualitative-quantitative data canyon," *American Journal of Health Studies*, vol. 16, no. 1, p. 52, 2000.
- [30] S. Abeyasekera, "Quantitative Analysis Approaches to Qualitative Data: Why, When and How? Methods in Development Research Combining qualitative and quantitative approaches," 2005, Statistical Services Centre, University of Reading, http://www.reading.ac.uk/ssc/workareas/participation/Quantitative_analysis_approaches_to_qualitative_data.pdf.
- [31] B. Simonetti, "An approach for the quantification of qualitative sensory variables using orthogonal polynomials," *Caribbean Journal of Mathematical and Computing Sciences*, vol. 13, pp. 16–24, 2006.
- [32] K. Bosch, *Elementare Einführung in die Angewandte Statistik*, Vieweg, 1982.
- [33] P. J. Zufiria and J. A. Berzal, "Analysis of hebbian models with lateral weight connections," in *Proceedings of the 9th International Work-Conference on Artificial Neural Networks*, vol. 4507 of *Lecture Notes in Computer Science*, pp. 39–46, 2007.
- [34] S. Müller and C. Supatgiat, "A quantitative optimization model for dynamic risk-based compliance management," *IBM Journal of Research and Development*, vol. 51, no. 3-4, pp. 295–307, 2007.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

