150330

Novel mathematical and statistical approaches to uncertainty evaluation

Case study of health care products

Deliverable 3.2.4

Work Package WP3: Decision-making and conformity assessment

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

1. Work package WP3	2. Deliverable number D3.2.4	3. Reporting date March 2015
4. Title Case study of health care products		
5. Author(s) SP, LGC, LNE, SCA		**6. Lead author (e-mail)** Leslie.Pendrill@sp.se
7. Contributing researchers (institutes)		
8. Other contributing work packages D3.1.5, D3.2.3		**9. Researchers in other WPs**
10. Supplementary notes		

11. Abstract

Best-practice in making multivariate conformity assessment and decision-making is illustrated in case studies of healthcare products. Mathematical and statistical approaches to uncertainty evaluation are introduced in studies where typically there is no simple theory of how the perceived ('response') properties **Z** depend on the physical material ('explanatory') properties **X** of the material. This is needed when dealing with properties important for consumers, such as smoothness of material surfaces (skin) as perceived in measurements by human panellists, which in turn depend on the surface topography, friction and hardness, of interest to the manufacturer wishing to fashion the product to the consumer's satisfaction. Principal component regression, consisting of an initial principal component analysis (PCA), can tackle correlation amongst the explanatory variables (**X**), thereby enabling both conformity assessment against explanatory variable **X** specification limits and multivariate regression including **Z**. A second challenge is that some measurements will be qualitative – such as the perceived 'slipperiness' in a current healthcare study – and require analysis on an ordinal scale, where many of the traditional tools of statistical analysis cannot be applied. Procedures for setting multivariate tolerance and specification regions and for calculating multivariate conformance properties are reviewed and applied. This includes validation of new software for bivariate conformance probability calculations developed in earlier parts of this NEW04 project. Methods for treatment of measurement uncertainty, reliability and decision risks in multivariate conformity assessment are developed, often drawing analogies with the treatment of corresponding product dispersion in traditional multivariate statistical process control.

12. Keywords: Measurement - Uncertainty - Multivariate – Tolerance – Specification - Decision theory – Conformity assessment - Risks – Costs - Healthcare

Contents

List of Tables 4
List of Figures 4
1 Case study of healthcare products 5
2 Software for conformity assessment in multivariate cases 5
3 Specific observations X assessment: healthcare study 6
4 Specific PC assessment: healthcare study 8
4.1 Specific PC Region 9
5 Global conformity assessment: conformance probabilities for bivariate cases – example health care products 11
5.1 Global observations X assessment: healthcare study 11
5.2 Global PC Region 13
6 Adding measurement uncertainty to calculation of conformance probabilities for bivariate cases: healthcare studies 13
6.1 X conformance probabilities in the healthcare study: global conformity, including measurement uncertainty 14
6.2 Allowing for correlation when including measurement uncertainty in conformance probability calculations 15
7 Multivariate regression 15
7.1 Dealing with qualitative responses 15
7.2 Results of PCR for healthcare studies 16
8 Conclusions 17
Acknowledgments 18
Annex A Conformance probabilities for multivariate cases 19
A.1 Multivariate analysis 19
A.2 Multivariate conformity assessment 19
A.2.2 Tolerance region providing a minimum coverage in manufacture 21
Annex B Including measurement uncertainty in bivariate conformance probabilities 24
B.1 Metrological (multivariate) coverage 24
B.3 Calculating multivariate conformance probabilities including measurement uncertainty 25
Annex C Multivariate regression 27
Annex D Rasch invariant theory and logistic regression 27

List of Tables

Table 3.1 Summary of calculated bivariate **X** conformance probabilities in the healthcare study: specific conformity……..7

Table 4.1 Summary of calculated bivariate PC conformance probabilities in the healthcare study: specific conformity……..11

Table 5.1 Summary of calculated bivariate **X** conformance probabilities in the healthcare study: global conformity……12

Table 5.2 Summary of calculated bivariate PC conformance probabilities in the healthcare study: global conformity……..13

Table 6.1 Summary of calculated bivariate PC conformance probabilities in the healthcare study: global conformity, including measurement uncertainty……14

Table 7.1 Summary of relations in the healthcare study between sensory response **Z** (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples and explanatory variables from PCR analyses……17

Table 7.2 Summary of calculated univariate sensory data **Z** conformance probabilities in the healthcare study: specific conformity, including measurement uncertainty, (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples and explanatory variables from PCR analyses……18

List of Figures

Figure 3.1 Correlation plot between a pair of explanatory variables X of six topical formulations [physical material properties: (x-axis) (friction) and (y-axis) (adhesion)] and corresponding specification limits for the healthcare case studies [Ringstad et al. 2015]…….6

Figure 4.1 Correlation plot between a pair of explanatory variables **X'** of six topical formulations [physical material properties: (x-axis) $\mu_{BS}FB_{0min}$ (friction) and (y-axis) $A_{BS}FB_{0min}$ (adhesion)] and PC corresponding specification limits for the healthcare case studies [Ringstad *et al.* 2015]. Dispersion intervals on each point are based on correlation matrix (measurement uncertainty neglected)……..10

Figures 7.1 Relations in the healthcare study between sensory response (Slipperiness) (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples, measured (y-axis) and predicted (x-axis) from explanatory variables (ForceBoard, BIOSKIN) given in eq. (7.1) and (7.2), respectively from PCR analyses……..17

Figure A.1 Geometry of tolerance and process regions [Abu Zahid & Sultana 2008]……22

Figure D.1 Separate measures of test person sensitivity (TP) and the slipperiness of the six specimen creams ……29

Figure D.2 Comparison of classical test theory (left) and Rasch (right) radar plots for four skin cream attributes and six specimens [Ringstad *et al.* 2015]……..30

1 Case study of healthcare products

Various healthcare studies [Prall 1973[1], Fall 2008[2], Ringstad *et al.* 2015] are chosen here to illustrate novel mathematical and statistical approaches to uncertainty evaluation, as reported in an intermediate report on case study on health care products in the earlier NEW04 deliverable D3.2.3. Such cases are characterised as dealing with properties important for consumers, such as smoothness of material surfaces (skin) as perceived in measurements by human panellists, which in turn depend on the surface topography, friction and hardness, of interest to the manufacturer wishing to fashion the product to the consumer's satisfaction.

These healthcare cases provide a number of challenges. Mostly, there is no simple theory of how the perceived ('response') properties **Z** depend on the physical material ('explanatory') properties **X** of the material, that is, how the expression $\mathbf{Z} = f(\mathbf{X})$ is to be formulated. Material and perceived quantities may be correlated in various ways and, in addition, limited measurement quality may enter on both sides of this equation.

A principal component regression [Annex A] can deal with this and consists of an initial principal component analysis (PCA) tackling correlation amongst the explanatory variables (**X**), followed by a **multivariate regression** including **Z**. A second challenge is that some measurements will be **qualitative** – such as the perceived 'slipperiness', but also some of the physiochemical properties such as material hardness, of a skin cream in the current healthcare study – and require analysis on an ordinal scale, where many of the traditional tools of statistical analysis cannot be applied [Annex D].

The main aim of the current report is to present best-practice in making multivariate conformity assessment and decision-making, illustrated in case studies of healthcare products, following an earlier review in the NEW04 deliverable D3.1.2.

2 Software for conformity assessment in multivariate cases

An account of conformance probabilities for multivariate cases is given in Annex A of this report. Conformance probabilities in multivariate cases needed in the present healthcare studies can be evaluated using available software, such as in CRAN-R [2013][3] for multivariate Normal and *t* distributions. MathCad contains a number of multivariate functions which we employ. Further, software has been developed as part of D3.1.5 of the present NEW04 project and is available specifically for calculation of conformance probabilities for bivariate cases. Four possibilities are offered in the D3.1.5 software, accessible through a main menu: "Raw data"; "Conditional distributions"; "Marginal distributions"; and "Joint probability distribution".

[1] J.K. Prall 1973 "Instrumental evaluation of the effects of cosmetic products on skin surfaces with particular reference to smoothness", *J. Soc. Cosmet. Chem.*, **24**: 693-707

[2] Andreas Fall 2008, "Towards Surface Softness by topographic modeling of haptic properties", *Master of Science Thesis in Biomaterials*, Department of Biological Physics, Chalmers university of technology, performed at SCA Hygiene Products AB, Göteborg, Sweden, 2008

[3] CRAN-R 2013, "Multivariate Normal and t Distributions: Computes multivariate Normal and *t* probabilities, quantiles, random deviates and densities", **Mvtnorm**, http://cran.r-project.org/web/packages/mvtnorm/

As we proceed with calculations of multivariate conformance probabilities in these healthcare studies, the target of each conformity assessment will be clearly specified: 'Global' conformity denotes the assessment of populations of typical entities, while 'specific' conformity assessment refers to inspection of single items or individuals, as defined by Rossi and Crenna [2006][4].

3 Specific observations X assessment: healthcare study

As a typical case to be studied in healthcare, Ringstad *et al.* [2015] have recently investigated how, for a set of topical formulations (skin creams), perception of various consumer-related quantities, **Z**, such as slipperiness and smoothness, when each cream is applied on human skin, is related to a number of physiochemical properties, **X**, measured when the same creams are applied on artificial skin substrates.

We introduce the study by making firstly a multivariate analysis of the measured physiochemical properties, **X**, such as $\mu_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (friction) and $A_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (adhesion), plotted in Figure 3.1 together with corresponding upper and lower specification limits (USL and LSL), respectively, as set by the manufacturer with due consideration to both what is feasible to manufacture but ultimately what is required by the consumer.

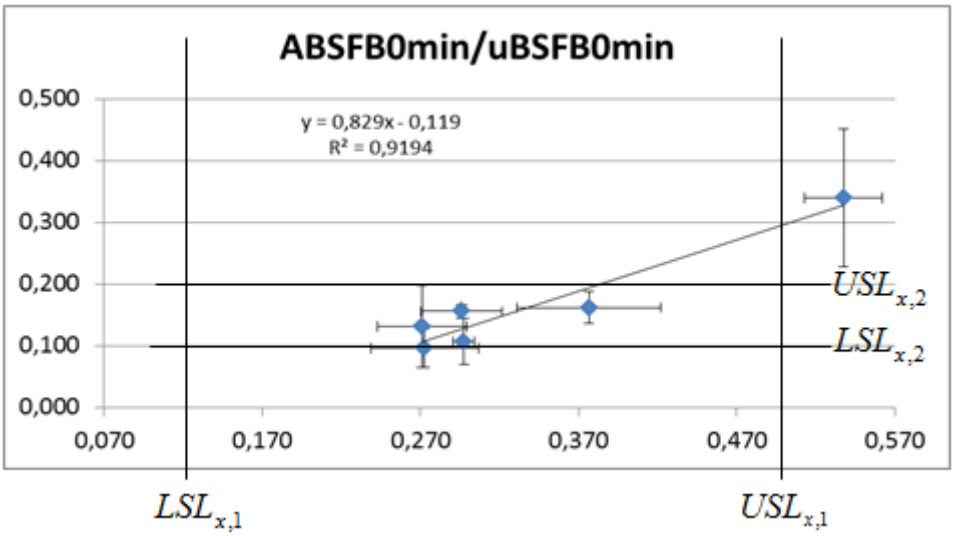


Figure 3.1 Correlation plot between a pair of explanatory variables **X** of six topical formulations [physical material properties: (x-axis) $\mu_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (friction) and (y-axis) $A_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (adhesion)] and corresponding specification limits for the healthcare case studies [Ringstad *et al.* 2015]

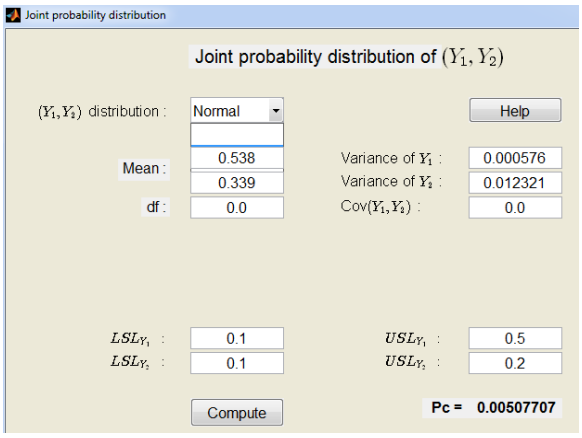
A initial study of <u>specific</u> conformity assessment - inspection of single items - was made in this bivariate healthcare case, where so-called 'consumer' and 'producer' percentage risks and conformance probabilities associated with measurement uncertainty, u_m, [Pendrill 2014a] were calculated using the software *MathCad*, assuming Normal distributions $N\left(x, u_m^2\right)$; no measurement

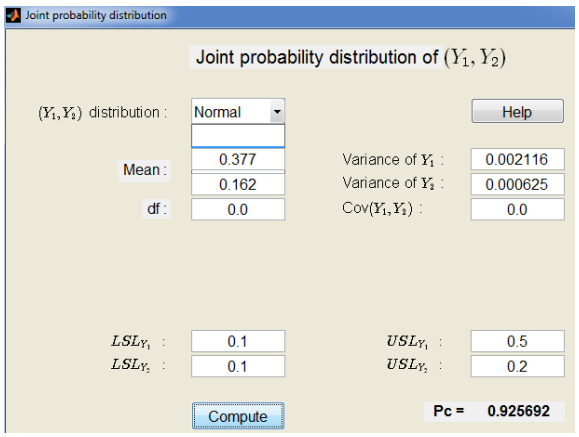
[4] G.B. Rossi and F. Crenna 2006, "A probabilistic approach to measurement-based decisions," *Measurement*, **39**, pp. 101 – 11

correlation; and no correlation between the two physiochemical properties **X** (i) $\mu_{BS}FB_{0min}$ (friction) and (ii) $A_{BS}FB_{0min}$ (adhesion), for two samples (B and F):

Table 3.1 Summary of calculated bivariate **X** conformance probabilities in the healthcare study: specific conformity

	Specific conformity assessment $(\hat{x}_i, u_m)$	Sample B $\mu_{BS}FB_{0min}$	Sample B $A_{BS}FB_{0min}$	Sample F $\mu_{BS}FB_{0min}$	Sample F $A_{BS}FB_{0min}$
		0,538(24)	0,34(11)	0,377(46)	0,162(25)
	USL	0,5	0,2	0,5	0,2
Consumer risk (accept/non-conforming)	$R_{M,P}(\hat{x}_i,u_m)=\Pr(x>USL\|\hat{x}_i,u_m)=\int_{USL_x}^{\infty}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}}\cdot dx;\hat{x}_i\le USL$			0,375%	6,426%
Producer risk (fail/conforming)	$R^*_{M,P}(\hat{x}_i,u_m)=\Pr(x\le USL\|\hat{x}_i,u_m)=\int_{-\infty}^{USL_x}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}}\cdot dx;\hat{x}_i>USL$	5,667%	10,524%		
	LSL	0,1	0,1	0,1	0,1
Consumer risk (accept/non-conforming)	$R_{M,P}(\hat{x}_i,u_m)=\Pr(x<LSL\|\hat{x}_i,u_m)=\int_{-\infty}^{LSL_x}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}}\cdot dx;\hat{x}_i\ge LSL$	0%	0%	1,565%	0,657%
Producer risk (fail/conforming)	$R^*_{M,P}(\hat{x}_i,u_m)=\Pr(x>USL\|\hat{x}_i,u_m)=\int_{LSL_x}^{\infty}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}}\cdot dx;\hat{x}_i\le LSL$				
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}_1,u_m)=\Pr(LSL_1<x<USL_1\|\hat{x}_1,u_m)=\int_{LSL_x}^{USL_x}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}}\cdot dx;LSL_1<\hat{x}_1<USL_1$	5,667%	8,959%	99,625%	92,918%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,M,P}(\hat{x}_1,u_m)\cdot P_{conf,2,M,P}(\hat{x}_1,u_m)$	0,5077066%		92,56923%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	0,507707%		92,5692%	





There is clearly excellent agreement from this validation study between bivariate conformance probabilities calculated with the MathCad and MatLab softwares.

4 Specific PC assessment: healthcare study

It is evident from Figure 3.1 that there is some degree of correlation amongst the pair of material property values plotted. This is not a measurement covariance but rather an actual correlation between the different material properties, e.g. one might expect there would be a relation between the adhesion and friction when specimens are formulated, for instance. Such correlation is not unusual, particularly in investigations such as the present where one deliberately chooses to measure a number of explanatory properties on which the consumer response is expected to depend in various ways.

Calculation of conformance probabilities will need to take account of this correlation, as will now be demonstrated with a principal component analysis [§A.2.2], as an example of PCA multivariate conformity assessment [§A.2.3] against principal component specification limits.

Covariance and correlation matrices, respectively, for the pair of explanatory variables **X** [physical material properties, $\mu_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (friction) and $A_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (adhesion) in the present healthcare case for a set of topical formulations (skin creams), plotted in Figure 3.1] were deduced from the experimental data with a MathCad program for PCA developed in the project:

$$\mathbf{Cov(X)} = \begin{bmatrix} 0{,}011 & 8{,}846\times10^{-3} \\ 8{,}846\times10^{-3} & 7{,}957\times10^{-3} \end{bmatrix};\ \mathbf{Cor(X)} = \begin{bmatrix} 1 & 0{,}959 \\ 0{,}959 & 1 \end{bmatrix} \tag{4.1}$$

using the MathCad function

$$\mathrm{Covar}(X) := \left| \begin{array}{l} \text{for } i \in 0..\,\mathrm{cols}(X) - 1 \\ \quad X^{\langle i \rangle} \leftarrow X^{\langle i \rangle} - \mathrm{mean}\left(X^{\langle i \rangle}\right) \\ \dfrac{X^{T}\cdot X}{\mathrm{rows}(X) - 1} \end{array} \right.$$

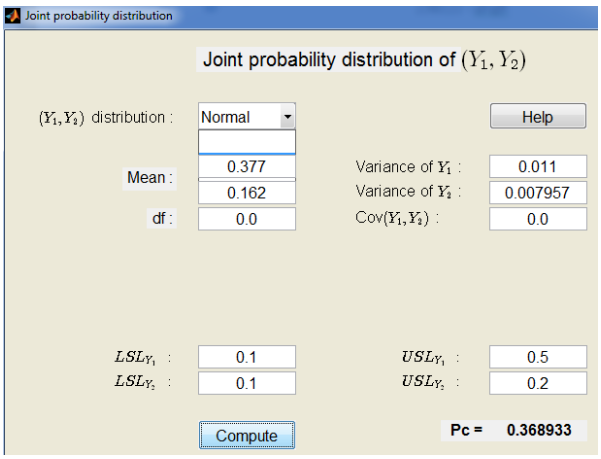
where the non-zero off-diagonal elements of **Cor(X)** indicate the degree of correlation.

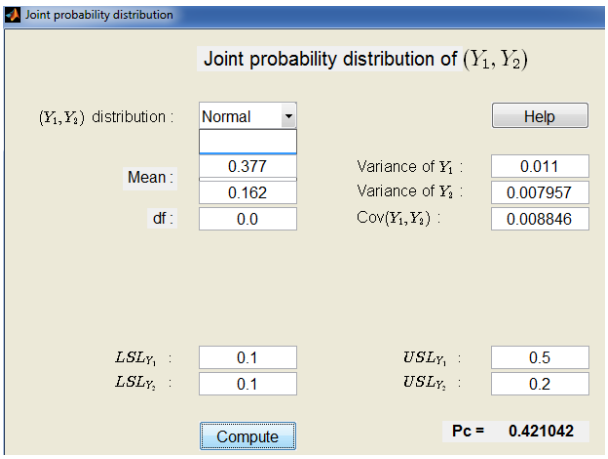
From Figure 3.1, it is clear that sample cream B had significantly different properties compared with the other 5 sample creams A, C – F. It is therefore somewhat of a challenge, with essentially only two distinct points amongst the six specimens, to provide reliable correlation estimates. The present healthcare study ideally would have to be extended to include additional specimen creams with other distinguishing ingredients, as studied recently for example by Lukic *et al.* [2012][5]. The corresponding 'goodness of fit' (R^2 in Figure 3.1) can be measured as the group variance, $\frac{1}{\mathrm{M-N}}\cdot\sum_i\left(z_{fit,i} - z_{\mathrm{i}}\right)^2$,

[5] M Lukic, I Jaksic, V Krstonosic, and S Savic 2012, "A combined approach in characterisation of an effective w/o hand cream: the influence of emollient on textural, sensorial and *in vivo* skin performance", *Intl. J. Cosmetic Science*, **34**, 140 – 9, doi: 10.1111/j.1468-2494.2011.00693.x

calculated from the regression residuals, $R_i = z_{fit,i} - z_i$, with a corresponding number of degrees of freedom equal to *M* - *N*, for the *M* (= 6) specimens amongst *N* (= 2) fitted parameters.

Bivariate conformance probabilities are calculated for each sample data of the pair of explanatory variables **X** and taking dispersion intervals on each point based on correlation matrix (while assuming no measurement uncertainty):





According to the new bivariate conformance probabilities Pc software (NEW04 D3.1.5), allowing for the covariances in **X** (from eq.(4.1)), conformance probabilities for one sample (e.g. item F) of 36,9% assuming zero covariance, increase to 42,1% including covariance.

4.1 Specific PC Region

To eliminate this covariance, which confounds the calculation of conformance probabilities, a PCA analysis [§A.2.2] in the current healthcare case is now made in which two (*k* = 2) principal components $\mathbf{p}_1$, $\mathbf{p}_2$ of variation amongst the physiochemical properties were identified:

$$\mathbf{P} = \begin{bmatrix} 0{,}759 & -0{,}651 \\ 0{,}651 & 0{,}759 \end{bmatrix} \tag{4.2}$$

using the MathCad function $\text{eigenvals}(\text{Covar}(X))$ to calculate the eigenvalues, λ_k, followed by $P^{\langle k \rangle} := \text{eigenvec}(\text{Covar}(X), \lambda_k)$ as an evaluation of eq. (A.7) .

The resulting covariance matrix:

$$\mathbf{Cov(X')} = \begin{bmatrix} 0{,}018 & 0 \\ 0 & 3{,}761 \times 10^{-4} \end{bmatrix} \tag{4.3}$$

for the transformed data set $\mathbf{X'} = \mathbf{X} \cdot \mathbf{P}$ clearly shows, with the zero off-diagonal elements, the orthogonality of the two PCs.

Adopting the approach of Wang and Chen [1998], a multivariate PC specification region $S_{PC1} = \{y_i \in \mathbf{X}' | \mathbf{LSL}_{PC} \le y_i \le \mathbf{USL}_{PC}\}$ [§A.2.4 (A.8)] in the current healthcare case is bounded by:

$$\mathbf{USL}_{pc} = \mathbf{p}^T \cdot \mathbf{USL} = \begin{bmatrix} 0{,}759 & 0{,}651 \\ -0{,}651 & 0{,}759 \end{bmatrix} \cdot \begin{bmatrix} 0{,}5 \\ 0{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}510 \\ -0{,}174 \end{bmatrix}$$

$$\mathbf{LSL}_{pc} = \mathbf{p}^T \cdot \mathbf{LSL} = \begin{bmatrix} 0{,}759 & 0{,}651 \\ -0{,}651 & 0{,}759 \end{bmatrix} \cdot \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix} = \begin{bmatrix} 0{,}141 \\ 0{,}011 \end{bmatrix}$$

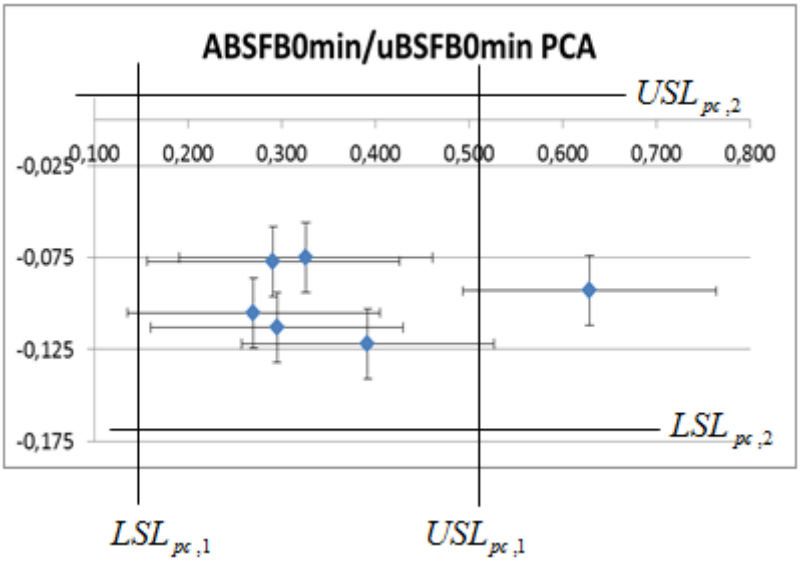


Figure 4.1 Correlation plot between a pair of explanatory variables **X'** of six topical formulations [physical material properties: (x-axis) $\mu_{BS}FB_{0min}$ (friction) and (y-axis) $A_{BS}FB_{0min}$ (adhesion)] and PC corresponding specification limits for the healthcare case studies [Ringstad *et al.* 2015]. Dispersion intervals on each point are based on correlation matrix (measurement uncertainty neglected)

Figure 4.1 indicates that most of the variability is concentrated in the first principal component, which of course reflects – as observed in Figure 3.1 – that the observations amongst the six samples are clustered in essentially only two distinct points, thus providing a challenge to reliable correlation estimation.

Applying the new PC specification limits to the modified data set $\mathbf{X}' = \mathbf{X} \cdot \mathbf{P}$, conformance probabilities can be calculated, assuming no measurement uncertainty in the pair of explanatory variables **X'** and taking dispersion intervals on each point based on correlation matrix (4.3):

Table 4.1 Summary of calculated bivariate PC conformance probabilities in the healthcare study: specific conformity

	Specific conformity assessment $(\hat{x}_i, u_m)$	Sample B $\mu_{BS}FB_{0min}$	Sample B $A_{BS}FB_{0min}$	Sample F $\mu_{BS}FB_{0min}$	Sample F $A_{BS}FB_{0min}$
		0,629(24)	-0,093(11)	0,392(46)	-0,122(25)
	USL	0,510	0,011	0,510	0,011
Consumer risk (accept/non-conforming)	$R_{M,P}(\hat{x}_i, u_m) = \Pr(x > USL \mid \hat{x}_i, u_m) = \int_{USL_x}^{\infty} \frac{1}{\sqrt{2\pi}\cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}} \cdot dx; \hat{x}_i \le USL$			19,114%	
Producer risk (fail/conforming)	$R^*_{M,P}(\hat{x}_i, u_m) = \Pr(x \le USL \mid \hat{x}_i, u_m) = \int_{-\infty}^{USL_x} \frac{1}{\sqrt{2\pi}\cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}} \cdot dx; \hat{x}_i > USL$	18,878%			
	LSL	0,141	-0,173	0,141	-0,173
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}_1, u_m) = \Pr(LSL_1 < x < USL_1 \mid \hat{x}_1, u_m) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi}\cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}} \cdot dx; LSL_1 < \hat{x}_1 < USL_1$	18,863%	99,998%	77,700%	99,584%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,M,P}(\hat{x}_1, u_m) \cdot P_{conf,2,M,P}(\hat{x}_1, u_m)$	18,863%		77,377%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	18,7409%		77,9757%	

Even in this PC assessment case, there is evidently excellent agreement from this validation study between bivariate conformance probabilities calculated with the MathCad and MatLab softwares.

The conformance probabilities of 78% for sample F calculated in the above table wrt the PC specification limits (*de facto* free of correlation) indicate an error of 47% in the initial estimate of Pc given as 36,9% given earlier in §4 for a comparable dispersion. This error is of the same order of magnitude as the predicted % error in Pc estimated with 95% correlation to be 42,6% for specification limits set at ± 3σ by Law [1996] who studied the effects of covariance on bivariate conformance probabilities. Subaric-Leitis [2010] has also studied the effects of covariance on conformance probabilities.

5 Global conformity assessment: conformance probabilities for bivariate cases – example health care products

A study of global conformity assessment, that is, the assessment of populations of typical entities, applied to the topical formulations (skin cream) example from the healthcare case studies (D3.2.3), gave another opportunity to evaluate the new bivariate conformance probability software developed as part of D3.1.5.

5.1 Global observations X assessment: healthcare study

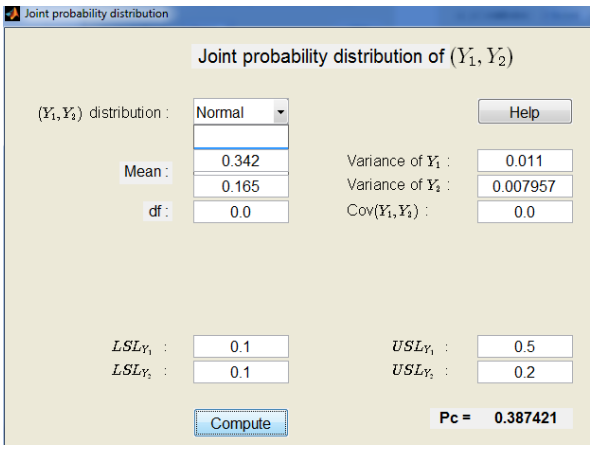
Global assessment in the present case means calculation of conformance probabilities for the typical values $\overline{\mathbf{X}} = \begin{bmatrix} 0{,}34(3) \\ 0{,}17(5) \end{bmatrix}$ from an average over the six formulation samples of the pair of explanatory

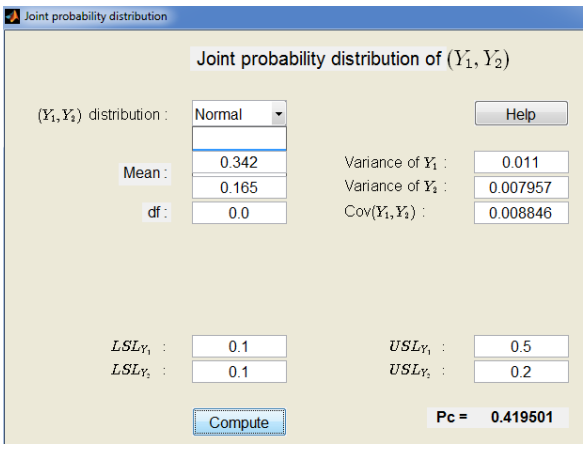
variables X [physical material properties $\mu_{VS}FB_{0min}$ (friction) and $A_{VS}FB_{0min}$ (adhesion) and measured with the *ForceBoard* set-up on artificial skin substrates *BioSkin*] with respect to specification limits $\mathbf{USL} = \begin{bmatrix} 0{,}5 \\ 0{,}2 \end{bmatrix}; \mathbf{LSL} = \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix}$, using the "Joint probability distribution" option with a *t*-distribution of $df = M - N = 4$ degrees of freedom ($M = 6$ samples and $N = 2$ explanatory variables). The corresponding multivariate specification region in this case is: $S_{\hat{X}} = \left\{ \hat{\mathbf{x}} \middle| \mathbf{LSL}_X \le \hat{\mathbf{x}} \le \mathbf{USL}_X \right\}$

Assuming no correlation and no measurement uncertainty in the pair of explanatory variables **X**:

Table 5.1 Summary of calculated bivariate X conformance probabilities in the healthcare study: global conformity

	Global conformity assessment $(\hat{x}, \sigma_P)$ σ_p calculated from diagonal elements of Cov(**X**)	Mean $\mu_{BS}FB_{0min}$ 0,342(103)	Mean $A_{BS}FB_{0min}$ 0,165(89)
	USL	0,5	0,2
	LSL	0,1	0,1
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}, \sigma_p(\hat{x})) = \Pr(LSL_1 < \hat{x} < USL_1 \mid \hat{x}, \sigma_p(\hat{x})) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi} \cdot \sigma_p(\hat{x})} \cdot e^{-\frac{(y-\hat{x})^2}{2 \cdot \sigma_p^2(\hat{x})}} \cdot dy; LSL_1 < \hat{x} < USL_1$	92,687%	41,964%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,M,P}(\hat{x}, \sigma_p(\hat{x})) \cdot P_{conf,2,M,P}(\hat{x}, \sigma_p(\hat{x}))$	38,895%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	38,7421%	





Accounting for the 0,959 coefficient of correlation (eq. 3.13) known from the PC analysis yielded instead a conformance probability (Normal distribution) of $P_{conform}$ = 41,9501%, illustrating the importance of accounting for correlation in such calculations.

5.2 Global PC Region

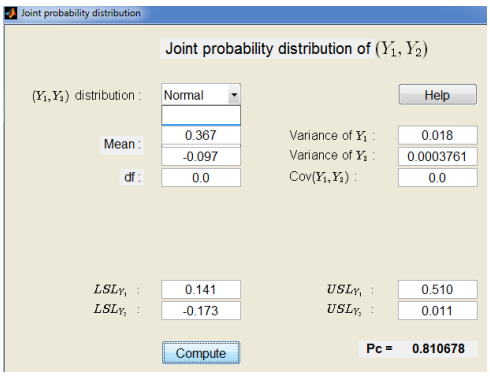
The new PC specification limits apply to the modified data set $\mathbf{X}' = \mathbf{X} \cdot \mathbf{P}$ where, again, conformance probabilities can be calculated for typical (mean) values

$$\overline{\mathbf{X}}' = \mathbf{p}^T \cdot \overline{\mathbf{X}} = \begin{bmatrix} 0{,}759 & 0{,}651 \\ -0{,}651 & 0{,}759 \end{bmatrix} \cdot \begin{bmatrix} 0{,}342 \\ 0{,}165 \end{bmatrix} = \begin{bmatrix} 0{,}367(135) \\ -0{,}097(19) \end{bmatrix}.$$

Assuming no measurement uncertainty in the pair of explanatory variables **X'**:

Table 5.2 Summary of calculated bivariate PC conformance probabilities in the healthcare study: global conformity

Global conformity assessment - PCA $(\mathbf{p}, \sigma_P)$ σ_p calculated from diagonal elements of Cov(**X'**)		Mean PC1 0,367(135)	Mean PC2 -0,097(19)
$\mathbf{USL}_{pc} = \mathbf{p}^T \cdot \mathbf{USL}$		0,510	0,011
$\mathbf{LSL}_{pc} = \mathbf{p}^T \cdot \mathbf{LSL}$		0,141	-0,173
Univariate conformance probabilities	$P_{conf,1,\mathrm{M,P}}(\hat{x}, \sigma_p(\hat{x})) = \Pr(LSL_1 < \hat{x} < USL_1 \mid \hat{x}, \sigma_p(\hat{x})) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi} \cdot \sigma_p(\hat{x})} \cdot e^{-\frac{(y-\hat{x})^2}{2 \cdot \sigma_p^2(\hat{x})}} \cdot dy;\ LSL_1 < \hat{x} < USL_1$	80,720%	99,996%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,\mathrm{M,P}}(\hat{x}, \sigma_p(\hat{x})) \cdot P_{conf,2,\mathrm{M,P}}(\hat{x}, \sigma_p(\hat{x}))$	80,716%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	81,068%	



These bivariate probabilities of global conformance for the mean are much higher in the PC analysis than in the corresponding observation matrix (X) analysis [§5.1], reflecting that the PC mean is more squarely centred on the specification region, as evident in a comparison of Figures 3.1 and 4.1.

6 Adding measurement uncertainty to calculation of conformance probabilities for bivariate cases: healthcare studies

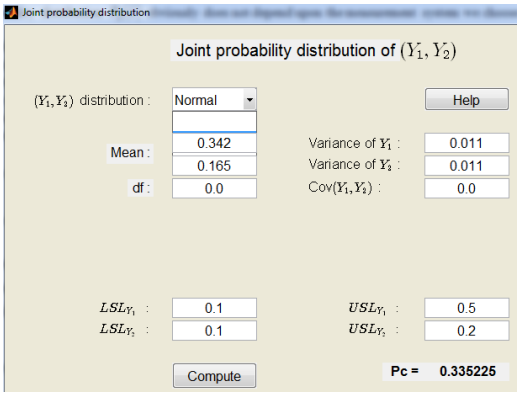
Traditional multivariate analysis has neglected measurement uncertainty. An account of how uncertainty can be included is given in Annex B.

6.1 X conformance probabilities in the healthcare study: global conformity, including measurement uncertainty

Assuming no correlation in the pair of explanatory variables **X** in the present healthcare case study:

Table 6.1 Summary of calculated **X** conformance probabilities in the healthcare study: global conformity, including measurement uncertainty but excluding correlations

	Global conformity assessment $(\hat{x}, \sigma_p(\hat{x}); u_m)$ σ_p calculated from diagonal elements of Cov(**X**) u_m, (B.3)	Mean $\mu_{BS}FB_{0min}$ 0,342(103;30)	Mean $A_{BS}FB_{0min}$ 0,165(89;57)
	Reliability coefficient, (B.1)	0,92	0,71
	USL	0,5	0,2
	LSL	0,1	0,1
	$\sqrt{\sigma_p^{\,2}(\hat{x}) + u^2{}_m}$	0,108	0,106
Univariate conformance probabilities	$P_{conf} = 100\% \cdot \left[1 - R_{\mathrm{M,P},U}(\hat{x}, \sigma_p(\hat{x}), u_m) - R_{\mathrm{M,P},L}(\hat{x}, \sigma_p(\hat{x}), u_m)\right]$ $R_{\mathrm{M,P},U}(\hat{x}, \sigma_p(\hat{x}), u_m) = \int_{-\infty}^{\infty}\int_{LSL_z}^{USL_z} \frac{1}{\sqrt{2\pi}\cdot u_m} \cdot e^{-\frac{(y-x)^2}{2\cdot(u_m^2)}} \cdot \frac{1}{\sqrt{2\pi}\cdot \sigma_p} \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot(\sigma^2{}_p)}} \cdot dy \cdot dx$	91,643%	35,939%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,\mathrm{M,P}}(\hat{x}, \sigma_p(\hat{x}); u_m) \cdot P_{conf,2,\mathrm{M,P}}(\hat{x}, \sigma_p(\hat{x}); u_m)$	32,935%	
Univariate conformance probabilities – Alt.	$R_{\mathrm{M,P},U,A}(\hat{x}, \sigma_p(\hat{x}), u_m) = \int_{LSL_z}^{USL_z} \frac{1}{\sqrt{2\pi}\cdot\sqrt{(\sigma^2{}_p + u_m^2)}} \cdot e^{-\frac{(y-\hat{x})^2}{2\cdot(\sigma^2{}_p + u_m^2)}} \cdot dy$	91,642%	35,939%
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	33,523%	



A value of R_β of 0,9 given in Table 6.1 means that about 75% of the observed variation is explained by product variations rather than limited measurement quality - which is considered acceptable [Linacre 2002][6]. Such limits, while enjoying widespread use, can appear to be somewhat arbitrary. Ultimately when appropriate (or fit-for-purpose) levels of uncertainty and associated risks of incorrect decisions

[6] J M Linacre 2002, "Optimizing Rating Scale Category Effectiveness", *J. Appl. Measurem.*, **3**, 85 - 106

are to be set, reference will need to be made to measures of impact for various stakeholder groups [Pendrill 2014a].

6.2 Allowing for correlation when including measurement uncertainty in conformance probability calculations

In looking for approaches to making in the best way a full account of limited measurement quality, including even correlation, in multivariate conformity assessment and decision-making, as elsewhere in the present work we can find useful analogies in the reasonably established treatment of multivariate entity dispersion.

Correlation in measurement – for example, arising from the use of a common instrument or measurement standard in different sets of measurements – could be dealt with in the multivariate context by employing a principal component analysis which would rotate the measurement dimension axes to new coordinates characterising the main variables of measurement dispersion. Note of course that this measurement PCA is conceptually distinct from the classical product PCA described in §A2.3, although of course the mathematics is similar.

7 Multivariate regression

The final part of present healthcare studies provide illustrative cases of the treatment of multivariable and qualitative observations where the way a product or material feels when touched is an important property for consumers.

A regression [Annex C] is to be made in order to find multivariate relations between the sensory response variables (**Z**) and various combinations of explanatory physiochemical variables (**X**) – an essential step which enables a product to be tailor-made by a manufacturer to suit consumer preferences, as expressed in terms of the PCR estimator, $\mathbf{Z} = \mathbf{p}^T \cdot \mathbf{Z}' = f(\mathbf{X})$.

The PC analysis performed above permits reduction of correlations amongst observed quantities **X** by identifying a set of orthogonal principal components (**P**), which in turn simplifies subsequent analyses, such as already demonstrated above for conformity assessment against explanatory variable **X** specification limits. This correlation freedom also enables a multivariate regression of the explanatory variables against response variables **Z**, as follows.

7.1 Dealing with qualitative responses

Since some measurements are **qualitative** – such as the perceived 'slipperiness' of a skin cream in the current healthcare study – and require analysis on an ordinal scale. To this end, we first introduce psychometric (Rasch) and generalised linear modelling[7] [Annex D] which to our knowledge have only recently entered[8] into the sensory community but which we consider an obligatory preliminary to any multivariate analysis of sensory data, traditionally performed with so-called quantitative descriptive

[7] L R Pendrill 2014b, "El ser humano como instrument de medida"/"Man as a Measurement Instrument", *e-medida* and *NCSLI Measure* J. Meas. Sci**., 9**, 24 - 35 (Dec 2014)

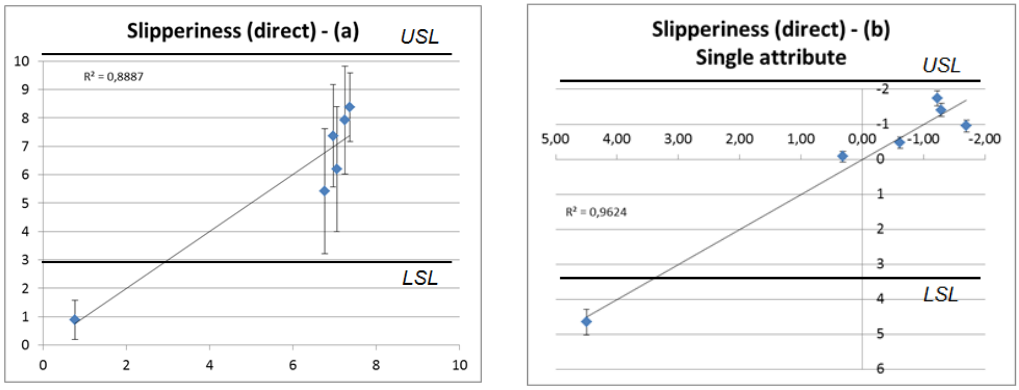
[8] F Marangon, L Pagani, G P Zaccomer, and S Troiano 2010, "Some uses of Rasch Model Parameters in Sensory Analysis: The example of Wine judging", Palerma 2010, http://www.academia.edu/8821621/Some_Uses_of_Rasch_Model_Parameters_in_Sensory_Analysis._The_Example_of_Wine_Judging

analysis (QDA) and other techniques[9]. Only when the raw data – e.g the sensory data **Z** – have been transformed to a quantitative scale, using the Rasch approach, can many of the 'traditional' tools of statistics and metrology then be applied without more ado[10]. This includes expressions of measurement reliability; measurement uncertainty and metrological traceability; precision and trueness; construct alleys; measurement/product separation; significance tests within and between raters, products, items, etc; spider plots; multivariate principal component analysis; etc.

Conceptually, this treatment of uncertainty in qualitative observations is intimately connected with decision risks [Pendrill 2014b].

7.2 Results of PCR for healthcare studies

In Figures 7.1 are shown for comparison typical results of principal component regressions (PCR) for (a) traditional test theory and (b) Rasch analysis [Annex D], providing explicit relations between sensory ('Slipperiness') and explanatory variables.



Figures 7.1 Relations in the healthcare study between sensory response (*Slipperiness*) (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples, measured (*y*-axis) and predicted (*x*-axis) from explanatory variables (*ForceBoard*, *BIOSKIN*) given in eq. (7.1) and (7.2), respectively from PCR analyses

It is evident from Figure 7.1 that the Rasch analysis of the sensory data yields lower measurement uncertainties than conventional test theory, thanks to Rasch analysis' ability to treat ordinal data properly as well as to provide separate estimates of item (sample cream) and person (sensory panellist) attribute values.

Table 7.1 Summary of relations in the healthcare study between sensory response **Z** (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples and explanatory variables from PCR analyses

	Slipperiness(direct), ForceBoard, BIOSKIN:	*USL*	*LSL*
(a) conventional test theory	$z_i = \text{Slipperiness}_i = 13{,}521 - 16{,}410 \cdot \mu_{BS}FB_{0min,i} - 11{,}364 \cdot A_{BS}FB_{0min,i}$ (7.1)	10	3
	$RMSE_{fit} = 0{,}75$		

[9] J M Murray, C M Dekahunty, and I A Baxter 2001, "Descriptive sensory analysis: past, present and future", *Food Research International*, **34**, 461 - 71

[10] E Svensson 2001, "Guidelines to statistical evaluation of data from rating scales and questionnaires", *J Rehab Med*; **33**: 47–48

(b) Rasch item parameters Θ)	$yR_i = SlipperinessR_i = -5{,}896 + 10{,}455 \cdot \mu BS_FB0min_i + 14{,}064 \cdot ABS_FB0min_i$ (7.2)	-3,4	2,1
	$RMSE_{fit}$ = 0,63		

The corresponding 'goodness of fit' ($RMSE_{fit}$ in Table 7.1) can be measured as the square root of the group variance, $\sqrt{\frac{1}{M-N} \cdot \sum_i (z_i - \text{Score}_i)^2}$, for each fit may be calculated from the regression residuals, $R_i = z_i - \text{Score}_i$, with a corresponding number of degrees of freedom equal to M - N, for the M specimens amongst N fitted parameters. In order to compare the relative merits of the various fits, compensation is made for the different scales of the conventional test theory and Rasch item parameters in calculating the group variances quoted in Table 7.1. The scaling factor is obtained from the slope of an approximate linear fit between the conventional test theory and Rasch sensory scores [Ringstad *et al.* 2015].

The sensory specification limits, *USL* and *LSL*, on the response variable **Z**, perceived slipperiness in the present healthcare study, given in Table 3.X correspond to specification limits on the pair of explanatory variables **X** [physical material properties $\mu_{VS}FB_{0min}$ (friction) and $A_{VS}FB_{0min}$ (adhesion) and measured with the *ForceBoard* set-up on artificial skin substrates *BioSkin*].

Table 7.2 Summary of calculated univariate sensory data **Z** conformance probabilities in the healthcare study: specific conformity, including measurement uncertainty, (a) conventional test theory, (b) Rasch item parameters Θ) for 2 samples (B & F) and explanatory variables from PCR analyses

	Specific conformity assessment Slipperiness, $(\hat{x}, u_m)$	Sample B (a)	Sample B (b)	Sample F (a)	Sample F (b)
		0,9(7)	5,4(2,2)	4,65(37)	-0,08(15)
	USL	10	-3,4	10	-3,4
	LSL	3	2,1	3	2,1
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}, u_m) = \Pr(LSL_1 < x < USL_1 \mid \hat{x}, u_m) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi} \cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2 \cdot u_m^2}} \cdot dx; LSL_1 < \hat{x} < USL_1$	0%	0%	100%	100%

8 Conclusions

Illustrations of best-practice in making multivariate conformity assessment and decision-making have been given in case studies of healthcare products, particularly a bivariate case of topical formulations (skin creams) where the final result was expressions for the relation between various consumer-related quantities, **Z**, such as slipperiness and smoothness, when each cream is applied on human skin, and a number of physiochemical properties, **X**, measured when the same creams are applied on artificial skin substrates.

This final result was achieved with a series of mathematical and statistical approaches to uncertainty evaluation. Procedures for setting multivariate tolerance and specification regions and for calculating multivariate conformance properties are reviewed and applied. This includes validation of new

software for bivariate conformance probability calculations developed in earlier parts of this NEW04 project. Principal component regression, consisting of an initial principal component analysis (PCA), tackling correlation amongst the explanatory variables (**X**), enabled both conformity assessment against explanatory variable **X** specification limits and multivariate regression including **Z**. A second challenge is that some measurements will be qualitative – such as the perceived 'slipperiness' in a current healthcare study – and require analysis on an ordinal scale, where many of the traditional tools of statistical analysis cannot be applied. Methods for treatment of measurement uncertainty, reliability and decision risks in multivariate conformity assessment are developed, often drawing analogies with the treatment of corresponding product dispersion in traditional multivariate statistical process control.

Acknowledgments

Thanks are due to other members of the NEW04 project for their constructive comments about the manuscript. The European Metrology Research Programme (EMRP, FP7 Art. 185) is jointly funded by the EMRP participating countries within EURAMET (www.euramet.org) and the European Union. The financial support and cooperation of SCA AB is also acknowledged.

Annex A Conformance probabilities for multivariate cases

A.1 Multivariate analysis

Table A.1 NEW04 Deliverables for multivariate cases

Deliverable number	Deliverable description	Lead Participant	Other Participants	Deliverable type	Delivery date
3.1.2	Report on conformity assessment for multivariate cases and conformity assessment involving cost function	SP	JV, LGC, LNE, PTB, NPL, SCA	Report	July 2013
3.1.5	Software for conformity assessment in multivariate cases developed	LNE	SP, LGC	Software	Jul 2014
3.2.3	Intermediate report on case study on health care products	SP	LGC, SCA	Report	Jul 2014
3.2.4	Case study on health care products completed and an article submitted to a trade journal	SP	LGC, SCA	Report, Article	Mar 2015

In a metrological study of conformity assessment in a multivariate situation, the clear resolution of the multivariability into components associated with production and measurement errors presents a challenge (at least as significant as in the univariate case [D3.1.1; Pendrill *et al.* 2013]). In traditional pragmatic approaches to multivariate analysis, measurement uncertainty is assumed to be negligibly small. As we look in the present healthcare studies to include explicitly measurement uncertainty in multivariate cases, useful analogies can be drawn with the reasonably established treatment of multivariate entity dispersion, as will be described below.

In the wealth of literature describing the area, multivariate analysis, in fields such as chemometrics [Olivieri *et al*. 2006[11]], has been described as the "efficient extraction of selective information from unselective data".

A.2 Multivariate conformity assessment

A specification region is bounded by (multivariate) specification limits set e.g. with reference to product specifications based on end-user requirements as well as technical limitations of manufacture. Estimates are then made of the fraction conforming and/or non-conforming product for the actual production (or process) dispersion compared with a specification region. In the first instance, the observed dispersion is dominated by actual variability in product and traditionally measurement dispersion has in most cases been assumed to be negligibly small.

[11] A C Olivieri, N M Faber, J Ferré, R Boqué, J H Kalivas and H Mark 2006, "Uncertainty estimation and figures of merit for multivariate calibration", *Pure Appl. Chem.*, **78**, 633 – 61

A.2.1 Specification region and conformance probabilities for observations X

An overall specification region for the observation matrix **X** in the multivariate case can be described for example as a 'hyperrectangle': $S_X = \{\mathbf{x} | \mathbf{LSL}_X \le \mathbf{x} \le \mathbf{USL}_X\}$ where lower respectively upper specification limits are set for **X** as follows in terms of the univariate specification limits:

$$\mathbf{LSL}_X{}^T = (LSL_1, ..., LSL_q); \qquad \mathbf{USL}_X{}^T = (USL_1, ..., USL_q)$$

There remains in the international literature a number of different approaches to defining this multivariate specification region, as will be discussed below.

An overall multivariate specification region has been proposed by Subaric-Leitis [2010][12] as the intersection of the specification regions $S_X = \{\mathbf{x} | \mathbf{LSL}_X \le \mathbf{x} \le \mathbf{USL}_X\}$ for the individual multiple variables (assumed independent) of the observation matrix **X**:

$$\bigcap_{i=1}^{q} \{X_i | \mathbf{LSL}_{X,i} \le \mathbf{X} \le \mathbf{USL}_{X,i}\} \tag{A.1}$$

The collective probability that all of the object variables lie in conformance within their multivariate tolerance region can be calculated with [Subaric-Leitis 2010]:

$$\Pr(\mathbf{LSL}_{X,1} \le X_1 \le \mathbf{USL}_{X,1} ... \mathbf{LSL}_{X,q} \le X_q \le \mathbf{USL}_{X,q}) = \int_{\mathbf{LSL}_{X,1}}^{\mathbf{USL}_{X,1}} ... \int_{\mathbf{LSL}_{X,q}}^{\mathbf{USL}_{X,q}} g(X_1, ..., X_q) dX_1 ... dX_q \tag{A.2}$$

When the different components are independent:

$$\Pr_{conform} = \prod_{i=1}^{q} \int_{\mathbf{LSL}_{X,i}}^{\mathbf{USL}_{X,i}} g(X_i, \overline{X}_i, u_i) dX_i \tag{A.3}$$

A multivariate process capability index wrt to the observations matrix **X** could in the case of indepenedent components be calculated in terms of the individual univariate indices:

$$C_{X,i} = \frac{USL_{X,i} - LSL_{X,i}}{6 \cdot \sigma_{X,i}} \tag{A.4}$$

where $\sigma_{X,i}$ is the process standard deviation for the ith variable of **X**, as given by the (square root of) the diagonal elements of the covariance matrix of the observations:

[12] A Subaric-Leitis 2010, "Risikoanalyse multivariater Konformitätsprüfungen". *tm - Technisches Messen*: **77**, pp. 662-670, http://dx.doi.org/10.1524/teme.2010.0091

$$\mathbf{Cov}[\mathbf{X}] = \begin{bmatrix} \sigma_1^{\,2} & r_{1,2}\cdot\sigma_1\cdot\sigma_2 & \dots & r_{1,q}\cdot\sigma_1\cdot\sigma_q \\ r_{2,1}\cdot\sigma_2\cdot\sigma_1 & \sigma_2^{\,2} & & \\ \dots & & & \\ \dots & & & \\ & & & \end{bmatrix}$$

A corresponding dimensionless correlation matrix is often used: $\mathbf{Cor}[\mathbf{X}] = \begin{bmatrix} 1 & r_{1,2} & \dots & r_{1,q} \\ r_{1,2} & & & \\ \dots & & & \\ \dots & & & \\ & & & \end{bmatrix}$,

involving only the correlation coefficients, $r_{i,j}$, and is obtained by dividing each element of **Cov[X]** by the product $\sigma_i \cdot \sigma_j$ of standard deviations of variability of pairs (i,j) of components.

In general, there will be some degree of correlation amongst the various observation variables, that is, the off-diagonal elements of **Cov(X)** are not zero. This complicates evaluation of conformance probabilities and process capabilities with equations (A.2) and (A.4), respectively.

A.2.2 Tolerance region providing a minimum coverage in manufacture

For completeness, mention is also made of multivariate tolerance regions.

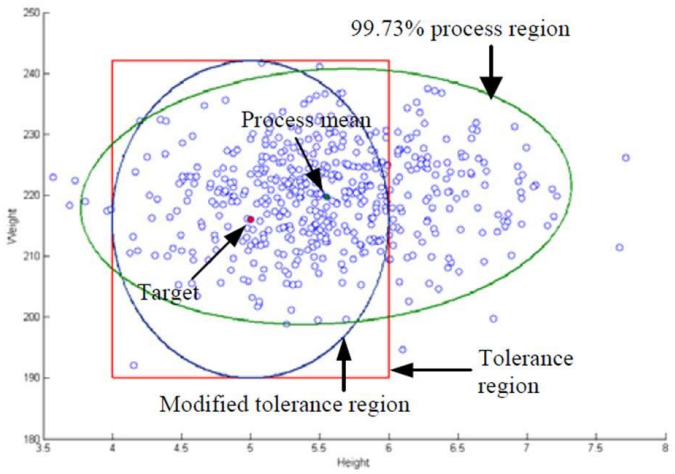


Figure A.1 - Geometry of tolerance and process regions [Abu Zahid & Sultana 2008][13]

[13] Md. Abu Zahid and A Sultana 2008, "Assessment and comparison of multivariate process capability indices in ceramic industry", *Journal of Mechanical Engineering*, **ME39**, 18 - 25

Instead of defining a specification region in terms of consumer-based specifications, a tolerance region (illustrated in Figure A.1) can be introduced by a manufacturer to provide a certain minimum coverage in production[14]. For instance, a $\frac{100\times(1-\alpha)\%}{100\times P\%}$ Normal tolerance region, gives, with $100\times(1-\alpha)\%$ confidence, a region where at least $100\times P\%$ of the population of entities (products) will be enclosed. Law [1996][15] describes setting a limit on the error in estimates of conformance probabilities by adjusting the multivariate tolerance region to limit the fraction, α, non-conforming items (which is equal to the area under the tail of the χ^2-distribution, with υ degrees of freedom):

$$\left(\mathbf{X}-\hat{\mathbf{X}}\right)^T\cdot Cov^{-1}(\mathbf{X})\cdot\left(\mathbf{X}-\hat{\mathbf{X}}\right)=\chi^2_{\nu,\alpha} \qquad \text{(A.5)}$$

The ellipsoid defined by (A.5) will include $100\times(1-\alpha)\%$ of the volume of the distribution.

It is important to make a clear distinction between the dispersion in entity (*P*) values from measurement dispersion (*M*) merely reflecting limited measurement quality [Annex B] – it is not immediately apparent from Figure A.1 how much of the observed dispersion is an artefact of limited measurement quality.

A.2.3 Specification region in Principal Component Analysis (PCA)

A main tool to handle multivariate situations – particularly in our case, decision-making and conformity assessment – is principal component analysis (PCA) on the observed data matrix for the explanatory variables (**X**) which "is a mathematical method of reorganising information in a data set of samples… to discover new variables 'Principal Components' (PCs), which account for the majority of the variability in the data. This enables us to describe the information with considerably fewer variables that was originally present." [Davies and Fearn 2004]. If there are many PCs identified, some selection may be made of those with the most loading.

For this PCA, a linear model is assumed with model errors $\boldsymbol{\varepsilon}_\mathbf{x}$:

$$\mathbf{X}=\mathbf{T}\cdot\mathbf{P}+\boldsymbol{\varepsilon}_\mathbf{x} \qquad \text{(A.6)}$$

and the data matrix (**X**) is first centred. The PCs are the eigenvectors of the covariance matrix of **X**, that is, the vectors $\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_k}$ satisfying the equations

$$\mathbf{Cov(X)}\cdot\mathbf{p_n}=\lambda_n\cdot\mathbf{p_n} \qquad \text{(A.7)}$$

where λ_n is the nth eigenvalue of $\mathbf{Cov}[\mathbf{X}]$.

[14] D S Young 2010 "tolerance: An R Package for Estimating Tolerance Intervals", *J. Stat. Software*, **36**, 1 - 39
[15] M J Law 1996, *Multivariate Statistical Analysis of Assembly Tolerance Specifications*, M.Sc.Thesis Department of Mechanical Engineering, BrighamYoung University

The variables **P** and **T**, defined by the relations (A.6) and (A.7), are referred to as *loadings* and *scores*, respectively.

PCA analysis permits reduction of correlations amongst observed quantities by identifying a set of orthogonal principal components. The freedom from correlation simplifies subsequent analyses, such as in the present healthcare studies when making (i) conformity assessment against explanatory variable **X** specification limits [§3] and (ii) multivariate regression of the explanatory variables against response variables **Z** [§7].

A.2.4 PC Specification Region: Approach of Wang and Chen [1998]

Engineering specifications for the PC_is and their target values in multivariate process control have been proposed in the approach of Wang and Chen [1998][16] as:

$$\mathbf{LSL_{PC_i}} = \mathbf{p_i^T} \cdot \mathbf{LSL}; \qquad \mathbf{USL_{PC_i}} = \mathbf{p_i^T} \cdot \mathbf{USL}; \qquad \mathbf{T_{PC_i}} = \mathbf{p_i^T} \cdot \mathbf{T} \tag{A.8}$$

The corresponding multivariate PC specification region in this approach is: $S_{PC1} = \{y_i \in \mathbf{X}' | \mathbf{LSL}_{PC} \le y_i \le \mathbf{USL}_{PC}\}$

In this context, a multivariate process capability index has been proposed by Wang and Chen [1998] as simply the product of the individual univariate principal components (*PC)* indices:

$$MC_p = \left[\prod_{i=1}^{k} C_{p;PC_i} \right]^{\frac{1}{k}} \tag{A.9}$$

where

$C_{p;PC_i} = \dfrac{USL_{PC_i} - LSL_{PC_i}}{6 \cdot \sigma_{PC_i}}$ is the univariate (principal component, PC_i) process capability index and where the first k PCs capture 90% of process variability.

[16] Wang FK and Chen JC 1998 ”Capability index using principal component analysis”, *Quality Engineering* **11**: 21 - 7

Annex B Including measurement uncertainty in bivariate conformance probabilities

Traditional multivariate analysis has to date mainly assumed that measurement uncertainty is negligibly small compared with actual entity dispersion. Among the JCGM Guides to the expression of uncertainty in measurement, the following are particularly relevant to the multivariate case: (i) the GUM [2008] itself; (ii) GUM-S2 [2011]: Extension to any number of output quantities; and (iii) GUM-S3 [to appear]: Developing and using measurement models. The current JCGM 106 document on conformity assessment addresses only the problem of univariate case.

B.1 Metrological (multivariate) coverage

A multivariate measurement model is one in which there is any number of output quantities [JCGM 102:2012 §3.8]

NOTE A matrix representation of the general form of multivariate measurement model is $\mathbf{h}(\mathbf{Y},\mathbf{X}) = \mathbf{0}$. $\mathbf{Y} = (Y_1,\ldots,Y_m)^T$ is a vector of *m* output quantities, *m* in number, in the multivariate measurement model, which constitutes the measurand, the quantity values of which are to be inferred from information about input quantities $\mathbf{X} = (X_1,\ldots,X_N)^T$ in the multivariate measurement model. $\mathbf{h} = (h_1,\ldots,h_m)^T$ and $\mathbf{Y}$ are matrices of dimension *m* x 1.

In metrology, a coverage region $\mathbf{R_Y}$ is specified as an *m*-dimensional space that contains $\mathbf{Y}$ with given coverage probability $\boldsymbol{p}$, given an estimate $\boldsymbol{y}$ of the output quantity $\mathbf{Y} = (Y_1,\ldots,Y_m)^T$ and its associated covariance matrix $\mathbf{U_y}$. In general, once the PDF for $\mathbf{Y}$ is established, for a region $\mathbf{R_Y}$, either the coverage probability for a specified coverage region or a coverage region for a specified coverage probability can be determined. Doing so is straightforward for a Gaussian PDF [see JCGM 102:2012 §§6.5.2, 6.5.3 and 6.5.4]. For other PDFs, a numerical method such as the Monte Carlo can be employed [adapted from JCGM 102:2012 §§6.5.1.2 and 6.5.1.4].

Particular coverage regions are [GUM-S2 §6.5]:

a) hyper-ellipsoidal coverage region centred at **y**, and
b) hyper-rectangular coverage region centred at **y**

An example of a coverage region for a vector output quantity is the 100*p*% hyper-ellipsoidal coverage region:

$$(\boldsymbol{\eta} - \mathbf{y})^T \cdot \mathbf{U}_\mathbf{y}^{-1} \cdot (\boldsymbol{\eta} - \mathbf{y}) = k_p^2 \qquad \text{(B.1)}$$

where $\mathbf{y}$ specifies its location; $\mathbf{U_y}$ specifies its shape; k_p specifies the size of the region; and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_M)^T$ is a variable describing possible values of the quantity $\mathbf{Y}$ [adapted from JCGM 102:2012 §7.7.2].

A metrological multivariate coverage region is analogous to (but conceptually distinct from) the production multivariate tolerance regions (A.10) discussed in Annex A.

B.3 Calculating multivariate conformance probabilities including measurement uncertainty

B.3.1 Reliability

Because of limited measurement quality, a measured item attribute value β can differ from the 'true' β', with a measurement error ε_β, $\beta = \beta' + \varepsilon_\beta$.

Measurement uncertainty in a test result – an apparent product dispersion arising from limited measurement quality – can be a concern in Conformity Assessment by inspection since, if not accounted for, uncertainty can:

- lead to incorrect estimates of the consequences of entity error
- increase the risk of making incorrect decisions, such as failing a conforming entity or passing a non-conforming entity when the test result is close to a tolerance limit.

It is therefore good practice to set proactively a limited on how large measurement uncertainty is allowed to be in relation to the product dispersion to be assessed. This measurement quality limit is often quantified in terms of corresponding reliability coefficient calculated as:

$$R_\beta = \frac{True\ variance}{Observed\ variance} = \frac{\mathrm{var}(\beta)}{\mathrm{var}(\beta')} = \frac{\mathrm{var}(\beta') - \mathrm{var}(\varepsilon_\beta)}{\mathrm{var}(\beta')} \qquad \text{(B.1)}$$

Practically in cases where measurement dispersion is of comparable size to actual product variation, it can be difficult to separate these [Rossi and Crenna 2006]. This will make evaluation of reliability coefficients according to (B.1) a challenge.

B.3.2 Combining production and measurement dispersions

Quoting Rossi and Crenna [2006]: "It is interesting to note that the (consumer) risk… depends both upon the quality of the measurement process, M, expressed by the characteristic function $P_{\mathrm{M}}(x|\hat{x})$, and the characteristics of the process P, summarised by the probability distribution $P_{\mathrm{P}}(\hat{x})$. These two characteristics combine in the following way: $P_{\mathrm{M,P}}(x,\hat{x}) = P_{\mathrm{M}}(x|\hat{x}) \cdot P_{\mathrm{P}}(\hat{x})$ since… the conditional probability $P_{\mathrm{M}}(x|\hat{x})$ does not depend on the (production) probability distribution of *x*, because it has been obtained solely on the basis of the characteristic of the measurement process, whilst the probability distribution $P_{\mathrm{P}}(\hat{x})$ obviously does not depend upon the measurement system we choose."

Pendrill [2007][17] gives examples in the univariate case of nesting of product and measurement integrals to treat conformance probabilities in conformity assessment without neglecting measurement uncertainty. For example, the 'consumer' risk in global conformity assessment, i.e. where the mean value appears to lie within the region of permissible values, R_{PV}, but dispersions lead to finite

[17] L R Pendrill 2007, "Optimised Measurement Uncertainty and Decision-Making in Conformity Assessment", *NCSLi Measure*, Vol. 2 No. 2 • June 2007, pp. 76 – 86

probabilities of the entity being out of tolerance, can be calculated with respect to both upper and lower specification limits:

$$R_{\mathrm{M,P},U}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)=\Pr\left(\hat{x}>USL\middle|\hat{x},\sigma_p(\hat{x}),u_m\right)=\int_{\hat{x}<USL}\int_{x>USL} g_{\mathrm{M,P}}(x|\hat{x})\cdot g_{\mathrm{M,P}}(\hat{x})\cdot dx\cdot d\hat{x}$$

$$R_{\mathrm{M,P},L}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)=\Pr\left(\hat{x}<LSL\middle|\hat{x},\sigma_p(\hat{x}),u_m\right)=\int_{\hat{x}>LSL}\int_{x<LSL} g_{\mathrm{M,P}}(x|\hat{x})\cdot g_{\mathrm{M,P}}(\hat{x})\cdot dx\cdot d\hat{x}$$

Postulating plausible Normal distributions for the measurement process $N_{\mathrm{M}}\left(\hat{x},u_m^2\right)$ and of the production process $N_{\mathrm{P}}\left(\hat{x},\sigma_p^2\right)$ of the entity of interest, for example the consumer risk wrt to the upper specification limit can be evaluated as:

$$R_{\mathrm{M,P},U}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)=\int_{-\infty}^{\infty}\int_{USL_z}^{\infty}\frac{1}{\sqrt{2\pi}\cdot u_m}\cdot e^{-\frac{(y-x)^2}{2\cdot u_m^2}}\frac{1}{\sqrt{2\pi}\cdot\sigma_p}\cdot e^{-\frac{(x-\hat{x})^2}{2\cdot\sigma_p{}^2}}\cdot dy\cdot dx \qquad \text{(B.2)}$$

where we assume that the production and process means coincide [Figure A.1].

Note again the need to keep a clear distinction between production and measurement contributions: actual dispersion in product indicates a certain rate of non-conforming items while the apparent dispersion due to limited measurement quality leads to a risk of incorrect acceptance of non-conforming product.

Corresponding expressions for conformance probabilities including both entity and measurement dispersions can then be simply derived for the consumer risk case $\hat{x}\in R_{PV}$ as:

$$P_{conf}=100\%\cdot\left[1-R_{\mathrm{M,P},U}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)-R_{\mathrm{M,P},L}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)\right]$$

An alternative expression for the conformance probabilities including measurement uncertainty is:

$$R_{\mathrm{M,P},U,A}\left(\hat{x},\sigma_p(\hat{x}),u_m\right)=\int_{LSL_z}^{USL_z}\frac{1}{\sqrt{2\pi}\cdot\sqrt{\left(\sigma^2{}_p+u_m^2\right)}}\cdot e^{-\frac{(y-\hat{x})^2}{2\cdot\left(\sigma^2{}_p+u_m^2\right)}}\cdot dy$$

In both expressions, measurement uncertainty in each of these calculations of global assessment was estimated by taking an average of the individual specific assessment uncertainties for the M (=6) samples:

$$u_m=\sqrt{\frac{\sum_i u^2{}_{m,i}}{M}} \qquad \text{(B.3)}$$

Annex C Multivariate regression

In a multivariate regression, the observed vector of outcomes (i.e., response variables **Z**) is regressed on the selected principal components (PCs) from the PC analysis as covariates (i.e. explanatory variables, **P**), using ordinary least squares regression (linear regression) to get a vector of estimated regression coefficients (with dimension equal to the number of selected principal components), that is in the form $\mathbf{Z}' = f(\mathbf{P})$.

For the basic linear model with errors, modelled as $\mathbf{Z} = \mathbf{T} \cdot \mathbf{C} + \boldsymbol{\varepsilon}_z$ accounting for dispersion in **Z,** an (unweighted) least-squares regression estimate of a design matrix is

$$\hat{\mathbf{C}} = \left(\mathbf{T}_0^{\mathrm{T}} \cdot \mathbf{T}_0\right)^{-1} \cdot \mathbf{T}_0^{\mathrm{T}} \cdot \mathbf{Z} \qquad \text{(C.1)}$$

for a 'calibration' data set $\mathbf{T_0}$, so that a predicted response is $\hat{\mathbf{Z}}_\mathbf{0} = \mathbf{T_0} \cdot \hat{\mathbf{C}}$. Here the principal component scores are $\mathbf{T}_{\mathbf{PC_i}} = \mathbf{p}_\mathbf{i}^{\mathbf{T}} \cdot \mathbf{T}$ [eq. (A.8)]. This is done in MathCad using the function $\mathrm{regress}(X1, Z, n)$ when fitting a polynomial of order *n* (=1 in the present study).

Finally, this vector $\mathbf{Z}'$ is transformed back to the scale of the actual covariates, using the selected PCA loadings (the eigenvectors **P** corresponding to the selected principal components) to get the final PCR estimator (with dimension equal to the total number of covariates), $\mathbf{Z} = \mathbf{p}^T \cdot \mathbf{Z}' = f(\mathbf{X})$ for estimating the regression coefficients characterizing the original model.

Annex D Rasch invariant theory and logistic regression

Ordinal data can be treated in terms of the probability, *P*, of successful decisions of classification of observed values into a number of discrete categories, $k = 1,\ldots, K$ of item properties when probed by a number of persons (or other probes), $i = 1,\ldots,N_{TP}$:

$$P(z) = \frac{e^z}{1+e^z} \qquad \text{(D.1)}$$

The binomial Rasch variant of (D.1) is where the 'latent' variable is expressed as the difference $z_{i,k} = \theta_i - \beta_k$, between pairs of attributes, θ and β, which can be assigned to any pair of components interacting, such (person, probe,…) and (material, task,…), respectively, of the system subject to assessment. This formulation is one of a family of so-called generalised linear models which provide links between the response (**Z**) and explanatory stimulus (**X**) variables covering not only explicitly non-linear responses but also poorly known responses to the explanatory variables. The approach is particularly advantageous when treating ordinal data in that it provides a linear, quantitative function of the difference in attributes:

$\log\left(\frac{P}{1-P}\right)_{i,k} = \theta_i - \beta_k$ as well as a clear separation of person and item scatter[18].

Measurement uncertainty, which is needed to make statements about apparent differences between values, such as the relative probabilities or quality scores of the successful sorting amongst different categories, can be dealt with in Rasch modelling as follows.

For dichotomous observations by person *i* of item *k*, the scored response

$$v_{i,k} = P_{i,k} \pm \sqrt{P_{i,k} \times \left(1 - P_{i,k}\right)} \qquad \text{(D.2)}$$

where $P_{i,k}$ is the probability of success, $P_{success}$, given in Eq. (D.1). The binomial error distributions for dichotomous scores approximate Gaussian distributions when accumulated across all the observations, as they are in the estimation process[19].

A gauge of measurement uncertainty in the estimated Rasch attribute values θ and β is expressed as a standard error, $SE(\theta_i, \beta_j) = \sqrt{\frac{1}{\hat{P}_{i,j} \times \left(1 - \hat{P}_{i,j}\right)}}$, derived from the Fisher information in the context of maximum likelihood estimation, for instance for a particular item *j*,

$$u(\varepsilon_\beta)_{all;j} = \text{realSE}(\beta)_{all;j} = \frac{1}{\sqrt{\sum_{i=1}^{TP}\left[\sum_{k=1}^{K} k^2 \times q_{i,j,k} - v^2{}_{i,j}\right]}}.$$

Better resolution and reliability in attribute results are achieved with a Rasch analysis compared with classical test theory, not affected by the abilities or attitudes of the particular persons measured, or by the difficulties or qualities of the particular sensory test items used to measure.

[18] L R Pendrill 2014b, "El ser humano como instrument de medida"/"Man as a Measurement Instrument", *e-medida* and *NCSLi Measure*, (in press Dec 2014)

[19] J. Linacre, "Rasch Model with an Error Term," *Rasch Measurement Transactions*, vol. 23, no. 4, p. 1238, 2010