BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
FACULTY OF CHEMICAL AND BIOENGINEERING

Analysis of attribute measurement systems

Booklet of thesis

submitted by

Emese Vágó
M.Sc. in Chemical engineer

Supervisor:

Dr. Sándor Kemény
Professor

Department of Chemical and Environmental Process Engineering

2011

1. Introduction

The aim of the measurement system analysis is to characterise the uncertainty of the measurement system and to decide if the gauge is capable to measure the given process parameter. The so-called gauge R&R analysis of variables measurement systems is a well-elaborated and widely used procedure. Its statistical background is the analysis of the variance components of the source of measurement uncertainty. In case of attribute measurement system the outcome of the measurement is not continuous but discrete variable, thus the analysis of variances cannot be applied, a different approach is needed.

The most referred literature sources on this field is the Automotive Industry Action Group's measurement systems analysis (AIAG MSA) Manual[1]. The method discussed here uses the reference value of the analysed parts. The reference value is a continuous characteristic behind the attribute type decision. Although the attribute gauge sorts parts into categories, in many cases the characteristic (reference value) may be measured (e.g. length, diameter, etc.) but the attribute gauge is used instead of it in practice due to cost and time or availability considerations. In my dissertation I have focused on the cases when the decision outcome is binary and the reference value is know during the gauge analysis.

2. Literature background, aims

The values of the continuous variable in the background of the attribute type decision can be divided into three regions. These are: accept, reject and gray zone. If a part is from the accept/reject zone it is always accepted/rejected. If a part comes from the grey zone, the decision is a random outcome. In the gray zone the probability of acceptance depends on the value of the continuous variable (reference value). The connection between the reference value and the probability of acceptance is represented by an S-shaped curve that is called "gauge performance curve" (GPC) in the literature. The GPC of an ideal attribute gauge is a step function, where the step is at the specification limit. It means that the parts below the specification limit are always rejected, the parts above the specification limit are always accepted (in case of lower specification limit). In reality the curve is not steep, but S-shaped and the location of the inflection point can be assigned as the location of step. If the location of the inflection point is not at the specification limit the gauge is biased. The location of the inflection point is that reference value, where the probability of acceptance is 50%.

---

[1] Automotive Industry Action Group. Measurement System Analysis; Reference Manual, 3rd ed. Detroit, MI: Automotive Industry Action Group, 2002; 125-140.

The measurement system analysis suggested in the AIAG Manual consists of two completely different approaches: the crosstabulation and the analytic method. For the crosstabulation method a random sample of 50 element (parts) is taken from the process. Three operators make decisions on each part with three repetitions. The reference value is measured thus the true quality of the parts is known. The Manual suggests the Cohen-kappa as a measure of agreement among the raters and also among the raters and the reference value. Relative frequencies are also suggested to measure the operator-reference value agreement: (i) the efficiency is the relative frequency of correct decisions (ii) the ratio of false acceptance and rejection is also calculated (iii) the repeatability is the relative frequency of those parts that were sorted into the same category with each repetition.

The AIAG manual suggests the use of the $P(bad|reject)$ and $P(good|accept)$ conditional probabilities. These can be estimated from the random sample with relative frequencies. $P(bad|reject)$ gives the ratio of bad parts among the rejected ones, $P(good|accept)$ gives the ratio of good parts among the accepted ones. It is a special characteristic of attribute gauges that these probabilities depend on the parts. More precisely they depend on the distribution of the reference value of parts. In the case when the distribution of parts changes (compared to that during the gauge analysis), the Manual suggests the use of the Bayes-theorem to calculate $P(bad|reject)$.

A sample of eight elements is needed for the AIAG Manuals' analytic method. The sample elements are selected so that they cover the gray zone of the GPC. The reference value of each part is measured and repeated observations are made on each part. (Only one operator is involved in this study.) The probability of acceptance is estimated for each part by the relative frequencies of acceptance in the sample. The reference values vs. the probabilities of acceptance are plotted on a kind of normal probability plot and a straight line is fitted to the point. It gives a graphical estimate of the GPC. The bias and repeatability of the gauge is read from this figure.

The repeatability, reproducibility and bias are clearly defined in the case of continuous measurement system analysis. Good repeatability means that if a given operator measures a given part with repetitions the results will be close to each other. The reproducibility is the measure of the operators' agreement. The perfect reproducibility means that the expected value of the measurements is the same for a given part, no matter which operator performs the measurement. The bias is the difference between the expected value of the measurement and the true value. In case of attribute measurement systems the definition and explanation of these terms is not that unambiguous.

The repeatability depends on the selected part. The crosstabulation and the analytic method define it in two completely different ways. If the reproducibility is wrong among two operators it means that one of them accepts/rejects the parts more often than the other. If the operator accepts or rejects the part more likely than the other could depend the measured part as well. In the crosstabulation method the reproducibility is related to the ratio of parts with different decisions. This ratio could be high not only because of the wrong reproducibility but because of the wrong repeatability as well. Thus this measure characterise the two properties together. The same is hold for the Cohen kappa. In the AIAG Manual the bias is defined as the distance between the inflection point of the GPC and the specification limit. According to other literature sources, the bias can be calculated from the random sample and it is the ratio of the false acceptance and the false rejection.

The literature uses different approaches to define repeatability, reproducibility and bias, there is no uniform method for the attribute gauge analysis. In my dissertation a critical discussion is presented of the existing methods of attribute gauge analysis (especially of the AIAG Manuals'). As a result several serious theoretical deficiencies are discovered. My aim was to suggest an attribute measurement system analysis method which overcomes these problems. Further requirements about the new method that it should describe the properties of gauge and characterise the measurement systems' capability.

3. Calculation methods

The literature suggests primarily the use of countdown indexes for attribute measurement system analysis. The method suggested in my dissertation is based on the logit model. The logit model belongs to the generalised linear model family, and it is fitted with the maximum likelihood method. The efficiency of the different methods were analysed partly analytically, partly with simulation.

Most of the statistical calculations in my dissertation were performed with the software STATISTICA 8.0[2]. For the simulation studies I have used Visual Basic script embedded into the program.

For the random effects approach of the measurement system analysis the fitting of a generalized linear mixed model was needed. I used the lme4[3] module of the open source R software for these calculations. The numerical integrations in my dissertation were performed with the MATHEMATICA[4] software.

[2] StatSoft, Inc. (2007). STATISTICA (data analysis software system), version 8.0. www.statsoft.com

[3] D. Bates, M. Maechler lme4 Version 0.999375-35

[4] Mathematica. version 6.0.0 (2007) Wolfram Research, Inc.

4. Results

My research work can be divided into two main parts. In the first part I have shown the flaws and deficiencies of the methods widely used for attribute measurement system analysis. In the second part of my work, I proposed a new method that overcomes the previously discussed problems.

4.1. Critique of the literature methods for attribute measurement system analysis

The AIAG Manuals' crosstabulation method suggests the use of Cohen kappa as a measure the agreement among raters. When calculating kappa, the decision outcomes on the same part are paired based on the order of repetition. As the repetitions are independent from each other this pairing is theoretically incorrect, and results in misleading kappa estimation.

The AIAG Manual suggests the use of Bayes-theorem to estimate the $P(bad|reject)$ and $P(good|accept)$ probabilities if the process is changed. The conditions of the application are not fulfilled, as the $P(reject|bad)$ and $P(accept|good)$ conditional probabilities on the right-hand side of the equation are not constant upon the process change. This follows from the fact that the probability of acceptance depends on the distribution of reference values of the parts. Thus if the distribution changes the $P(reject|bad)$ and $P(accept|good)$ probabilities change as well.

The crosstabulation method of the AIAG Manual distinguish only the accept and the reject decision. However both the upper and the lower specification limits are given in the discussed example, thus the "too small" and "too large" parts can be distinguished among bad parts. These two possible outcomes are merged to the "reject" category. This simplification leads to loss of information. In the discussed example there are really not two but three outcome categories, and in these cases different approach should be used for the analysis.

Within the framework of crosstabulation method, different probabilities are estimated with countdown indexes. The estimation is based on the results of the three times repeated observations on a sample with 50 elements. The estimation would be more efficient, if larger sample (150 in this example) were used without repetitions.

The AIAG Manual suggests two completely separated methods for the measurement system analysis without any connection among them. There is a characteristic feature (repeatability) that is defined in two different ways with the two approaches.

The critiques and suggestions about the literature are summarised in ***theses 1*** and ***2***.

4.2. Modell-based approach for attribute measurement system analysis

A new method is proposed for the analysis of attribute measurement system analysis. The basic idea of the new method is the mathematical modelling of the gauge performance curve instead of the previously applied graphical interpretation. The proposed model uses the logit transformation to linearise the connection between the probability of acceptance and the reference value.

$$\text{logit}(p_i) = \ln\frac{p_i}{1-p_i} = \alpha + \beta_i^O + \beta^P x + \gamma_i^{O*P} x$$

where p_i is the probability that operator i accepts a part having x reference value

x value

α model constant

β_i^O effect of the i-th operator

β^P effect of the “part”

γ_i^{O*P} operator-part “interaction” at the i-th operator

Using the model above the following properties of the curve can be calculated:

- the bias of the gauge, i.e. the distance between the inflection point and the specification limit;

- the width of gray zone: an x-range in which the probability of acceptance is between ε and 1-ε, where ε is an arbitrarily small probability e.g. 0.05. In other words the S-shaped part of the curve.

The estimation of the model parameters is performed using a sample with known reference values. The sample elements are analysed by different operators with repetitions. Maximum likelihood method is used to estimate the model parameters. I have investigated if the efficiency of the estimation can be improved with the application of ridge method. The ridge estimators are usually applied in order to decrease the high variance due to the correlation of the independent variables. The independent variables are not correlated when estimating the parameters of the GPC, but due to the form of the logit model it can be assumed that the ridge method improves the estimate. I have investigated this hypothesis with simulations. The results are in line with the expectation, the error of the logit model parameters were substantially smaller with the ridge estimation in the simulation study. This improvement is much smaller in the error of the estimation of the inflection point and the gray zone. As the ridge method does not improve the results of practical importance, its application is not justified.

The attribute gauge can be perfectly characterised with the location of the inflection point and the width of gray zone. However these characteristics are only about the measurement system, and do not provide any information about the probability of making wrong decisions when using the gauge, which latter is the primary interest of the user. I proposed the use of the $P(good|reject)$ and $P(bad|accept)$ conditional probabilities as a measure of capability, as these conditional probabilities are connected directly to the loss caused by a wrong decision. If the distribution of the reference values is known these probabilities can be calculated using the mathematical model of the GPC. For the *i*-th operator:

$$P(good|reject)_i = \frac{\int_{x=SL}^{x=+\infty} f(x)[1-p_i(x)]dx}{\int_{x=0}^{x=+\infty} f(x)[1-p_i(x)]dx}$$

$$P(bad|accept)_i = \frac{\int_{x=0}^{x=SL} f(x)[p_i(x)]dx}{\int_{x=0}^{x=+\infty} f(x)[p_i(x)]dx}$$

Where $f(x)$ denotes the distribution of the reference value and *SL* is the specification limit.

Great advantage of the proposed calculation method that the characterisation of the gauge and the distribution of reference value are fully separated. This solves the problem of the AIAG Manuals' calculation based on the Bayes-theorem. The model-based calculation is valid even if the process has changed compared to those when estimating the GPC parameters.

The $P(good|reject)$ and $P(bad|accept)$ conditional probabilities can be estimated without fitting the model based on relative frequencies following the AIAG Manuals' crosstabulation method. I studied the efficiency of the two estimation method partly analytically, partly with simulation. The mean square error of the estimation is dramatically smaller with the proposed model-based method in the whole range of investigated sample sizes. Furthermore the error of the model-based estimation depends only slightly on the number observations above a practically feasible number of observations. In the lower range of the investigated sample sizes the AIAG-type calculation provides zero estimated probabilities with large probability. In these cases the performance of the AIAG-type estimation is especially poor compared to the model-based estimation.

In case of continuous measurement system the uncertainty of the outcome has two sources: the difference among the parts and the uncertainty of the measurement system. This latter also has two main parts: the uncertainty due to operators (reproducibility) and the uncertainty independent from operators (repeatability). Each source of variability is characterised by a variance component. As the parameters of the random effects model are variances, this approach can be used to characterise the attribute measurement systems analogously to the continuous measurement systems. I have investigated with simulations how $P(good|reject)$ and $P(bad|accept)$ can be estimated with the random effects approach.

In the first case (random intercept model) only the position of the GPC was different from operator to operator, the location of gray zone was the same. I have estimated the expected value and variance of $P(good|reject)$ and $P(bad|accept)$ with respect to the operators both with mixed and with fixed model. (The mixed model uses the random effects concept.) The error of the logit model parameters were smaller with the mixed model, but this difference were not present in each case in the estimates of the expected value and variance of $P(good|reject)$ and $P(bad|accept)$. In case of the fixed model I have estimated the expected value and variance of $P(good|reject)$ and $P(bad|accept)$ in two different ways: in a simpler way using mean and sample variance and with a more complex calculation using integrals. The error of the latter approach is somewhat smaller but the difference is minor.

In the second model (random intercept and random slope) both the position and the shape of the GPC vary randomly between operators. Here only the simpler fixed effect model was fitted. The simulation results show that the error of the more complicated estimation is smaller, but the difference is minor similar to the random intercept case.

As a final conclusion the estimates of the expected value and variance of $P(good|reject)$ and $P(bad|accept)$ are of proper order of magnitude with all estimation methods. The fitting of the mixed model is easier in practice while its software implementation is more widely solved (compared to the mixed model). Therefore, I suggest the use of the fixed model. I also suggest the use of the simpler calculation method (sample mean and variance) for the estimation of the expected value and variance of $P(good|reject)$ and $P(bad|accept)$, as the improvement achieved with the integration does not justify the additional difficulties of this calculation. In case of the random intercept and random slope model only the simpler fixed effect model was fitted. I suggest the use of the simpler calculation method for the estimation of the expected value and variance of $P(good|reject)$ and $P(bad|accept)$ on the same ground as in the random intercept case.

The results of the model-based measurement system analysis are summarised in ***theses 3, 4*** and ***5***.

5. Theses

Thesis 1

I have revealed the following theoretical errors and deficiencies of the attribute measurement system analysis methods proposed in the widely used AIAG Manual[2].

(a) The typically assigned 50 sample items with three decisions are treated as if they were 150 items. This improper handling of data leads to misleading estimation of kappa.

(b) If the distribution of the reference value of parts changes the probability of acceptance of the good and the bad parts changes as well. From this follows that the $P(reject|bad)$ and $P(accept|good)$ probabilities of the original process can not be applied when estimating $P(bad|reject)$ and $P(good|accept)$ probabilities with the Bayes-theorem. Thus the Bayes theorem is improperly used.

(c) The guideline uses two distinct methods: the crosstabulation and the analytical methods. Two different samples are needed for the two methods, their mathematical analysis is fully separated even if they are intended to characterize the same system. A conclusion of this separation is for example that two different definitions are given to the same term, repeatability.

Thesis 2

I found and justified that the phenomenon of overdispersion for binomial and multinomial distribution occurred only in cases of repeated assignments [4]. The consequence of this fact referring to the attribute measurement system analysis is that when using countdown indexes the estimation is more efficient if larger samples are assigned once than if smaller samples assigned with repetitions, assuming that the total number of assignments is the same.

Thesis 3

Instead of the generally used AIAG method I have proposed a new approach for attribute measurement system analysis [3]. The basic novelty of the new method is that it uses the mathematical model of the gauge-performance-curve. The dependent variable of the model is the logit transformed probability of acceptance, the independent terms are the reference value of part, and the operator.

(a) I have suggested the use of two characteristics of the gauge-performance-curve. These are the width of gray zone and the bias. I showed how they can be calculated using the proposed model.

(b) The model based approach allows more efficient estimation of the probability of wrong decisions as compared to the AIAG's crosstabulation approach.

(c) The novel method separates the characterization of the distribution of the reference values from the characterization of the gauge. The advantage of this separation is that if the manufacturing process is changed the gauge characterization need not be repeated, only the distribution parameters of the reference values need to be estimated from a new sample.

Thesis 4

I studied the application of the ridge method [1] to the estimation of the gauge performance curve and investigated the estimation method with simulations. The use of the ridge method decreased the estimation error of the logit model parameters in a great extent. However this improvement is much smaller in the error of the estimation of the inflection point and the gray zone. As the ridge method does not improve the estimation of the gauge characteristics of practical importance, its application is not justified.

Thesis 5

I have applied the random effects approach to the model-based measurement system analysis [5]. I have investigated two kinds of mixed models: in the first case (random intercept model) only the position of the GPC was different from operator to operator, in the second case (random intercept and random slope) both the position and the shape of the GPC vary randomly between operators. The expected value and variance of wrong decisions with respect to the operators can be estimated in both cases. Different approaches (different simplifications in calculations) can be applied during the estimation. I have investigated the effect of these simplifications to the estimation error with simulation.

The estimates are of proper order of magnitude with each approach. Therefore, the use of the simplest calculation method can be accepted for practice.

6. Possible practical applications

The attribute measurement systems are widely used in different fields of industry. There is often a continuous variable in the background of the decision. Therefore the analysis of this kind of measurement systems is of practical interest.

In my dissertation I have showed that the most widely used methods of attribute measurement system analysis discussed in the AIAG Manual are flawed in several aspects. Considering further literatures on the field I have found that the interpretation of the applied characteristics is not uniform. The different approaches for attribute measurement system analysis answer only a part of the questions that are of practical interest. The most important points are the characterisation of the properties of the gauge performance curve, the difference between operators and the capability of the gauge. None of the analysis methods discussed in the literature can answer each question.

I have proposed a new model based method approach for attribute measurement system analysis that is free from the theoretical errors of the AIAG Manuals' approach. Each question of practical interest related to the measurement system analysis can be answered with the proposed method. Further advantage of the new method that it needs fewer measurements thus its practical application will spread hopefully.

7. Publications

Publications of the theses

Papers in English

1. **E. Vágó**, S. Kemény: Logistic ridge regression for clinical data analysis (a case study). *Applied Ecology and Environmental Research* 4(2), 2006.

2. **E. Vágó**, S. Kemény: Critique of the AIAG Cross-Tabulation Procedure for Attribute Gauge R&R Study, *International Journal of Quality Engineering and Technology*, 2(1), p. 75-93, 2011, DOI: 10.1504/IJQET.2011.038724

3. **E. Vágó**, S. Kemény: A model based approach for attrinute R&R Analysis, *Quality and Reliability Engineering International*, 2010, DOI: 10.1002/qre.1154

4. **E. Vágó**, Zs. Lang, S. Kemény: Overdispersion at the Binomial and Multinomial Distribution, *Periodica Polytechnica* 55(1), p.17-20, 2011, DOI:10.3311/pp.ch.2011-1.03

5. **E. Vágó**, S. Kemény: Random effects model for attribute gauge R&R *Quality and Reliability Engineering International*, (accepted)

Presentations and posters on international conferences

1. **E. Vágó**, S. Kemény: Analysis of attribute measurement systems. *Conferentia Chemometrica* VII, Hajdúszoboszló (2005)

2. **E. Vágó**, S. Kemény: Analysis of attribute measurement systems. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry* (*SCAC*), Tihany (2006)

3. **E. Vágó**, S. Kemény: Overdispersion for multinomial distribution: a Case Study for Attribute Gage Analysis. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry* (*SCAC*), Balatonalmádi (2008)

Presentations on Hungarian conferences or scientific forums (in Hungarian)

1. **E. Vágó**, S. Kemény: Analysis of attribute measurement systems (Minősítéses mérőeszközök képességvizsgálata) *Műszaki Kémiai Napok*, Veszprém (2005)

2. **E. Vágó**, S. Kemény: Qualification of qualification (A minősítés minősítése). *Műszaki Kémiai Napok*, Veszprém (2006)

3. **E. Vágó**, S. Kemény: Overdispersion. *Klinikai Biostatisztikai Társaság ülése*, Budapest (2008)

Other publications connected to the PhD research work

Co-author of book

Kemény Sándor - Deák András - Lakné Komka Kinga - **Vágó Emese**: Statistical analysis by the Statistica software, (Statisztikai elemzés a STATISTICA programmal, in Hungarian), Műegyetemi Kiadó, Budapest, 2004

Oral lectures, posters

1. **E. Vágó**: Paired t-test and analysis of variance (Páros t-próba és varianciaanalízis in Hungarian). Műszaki Kémiai Napok, Veszprém (2002)

2. **E. Vágó**: Paired t-test and analysis of variance (Páros t-próba és varianciaanalízis in Hungarian). *EOQ Six Sigma Szakbizottság ülése*, Budapest (2003)

3. **E. Vágó**: Application of ridge regression for logit model (Ridge regresszió alkalmazása logit modell esetén in Hungarian). BME, *Vegyész és Biomérnöki Kar Doktoráns Konferencia*, Budapest (2004)

4. **E. Vágó**, S. Kemény: Paired t-test and analysis of variance. *Modern Methods of Data Analysis, Russia*, Belokurikha, (2003)

5. **E. Vágó**, S. Kemény: Ridge regression method to logit model. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry* (*SCAC*), Balatonfüred (2004)

6. **E. Vágó**, Kemény Sándor: Logistic ridge regression for clinical data analysis (a case study) *VII. Magyar Biometriai és Biomatematikai Konferencia*, Budapest (2005)

7. **E. Vágó**, S. Kemény: Application of ridge regression for logit model (Ridge regresszió alkalmazása logit modell esetén in Hungarian). *Klinikai Biostatisztikai Társaság ülése*, Budapest (2005)