# Definition of a Software Product Line Portfolio Using the Kano Model

*Karen Cortes Verdin[1,2,*] , Lorena Lopez Lozada[1]*
*[1]School of Statistics and Informatics*
*Universidad Veracruzana*
*Av. Xalapa esq. Avila Camacho S/N*
*Xalapa, Ver. 91020, Mexico*
*[2]Computer Science Department*
*Center for Research in Mathematics (CIMAT, A.C.)*
*Calle Jalisco S/N*
*Col. Mineral de Valenciana Guanajuato, Gto. 36240, Mexico*
*karen@cimat.mx, lorlopez@uv.mx*

## Abstract

*Software Product Line Engineering (SPLE) attempts to capitalize reuse by setting a framework for planning, development and management of core assets. Product Management is the "Planning, organizing, executing, and controlling of all tasks, which aim at a successful conception, production, and marketing of all products offered by a company". For SPLE, the definition of products entails great importance since such definition lays the foundations for reuse. The Kano Model provides a scheme for defining products that will yield high customer satisfaction. The model, however, lacks of statistical elements and principles. Therefore, with the involvement of students of Statistical Consulting a complete process for the use of the Kano Model could be obtained. Here, the use of the Kano Model for the definition of a Software Product Line product portfolio is presented.*

## 1. Introduction

A Software Product Line (SPL) is a "set of software-intensive systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way" [1]. SPLE is based on planned and proactive reuse, architecture-centric development and a two-tier software development organization.

SPLE encompasses product portfolio definition (also known as scope definition). A product portfolio is the set of product types offered by the Product Line (PL) organization. In product portfolio management new projects are evaluated, selected, and prioritized, while existing projects may be accelerated, cancelled, or de-prioritized. It has been demonstrated that companies with product management process obtain higher profitability that the industry average [2]. Besides, product portfolio management can also lay the foundations for core assets definition and reuse by identifying product interdependencies and variants. Core assets are the basis for product production, they are designed to be reused and optimized in more than one system [1]. Core assets include reusable software components, the architecture, domain models, requirements, performance models, documentation and test cases.

The Kano Model [3] allows the definition of products that will yield high customer satisfaction. The model is based on a classification scheme of customer requirements.

This paper presents the use of this model for the definition of a PL portfolio. This work was performed in collaboration with the students of the Statistical Consulting I course of the B. Sc. degree in Statistics at the School of Statistics and Informatics of the Universidad Veracruzana. The purpose of this course is to prepare students to be business leaders by developing their capabilities to use statistical thinking in business processes. Currently, statistical education suggests that students get involved in real projects of other disciplines [4] [5]. This paper is organized as

[*] Ph. D. Candidate at CIMAT, A.C.

follows: Section 2 presents a description of the Kano Model; Section 3 explains the process defined along with the case study used for such purpose. Finally, conclusions are given.

## 2. Kano Model

The Kano Model is a theory developed in the 80's by Professor Noriaki Kano [2] [3]. This is a theory of product development which focuses on differentiating product features, as opposed to focusing initially on customer needs. Kano Model provides a classification of product's requirements or characteristics based on how they are perceived by the customer and their effect on customer satisfaction. This classification can be observed in Figure 2.1 below.

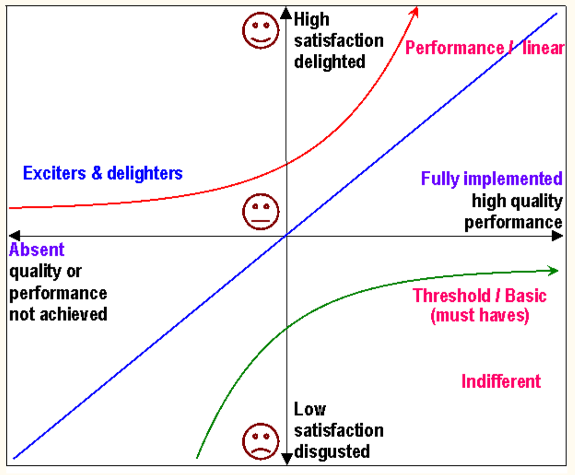

**Figure 2.1 Requirements classification according to Kano Model.**

The categories of the Kano Model classification scheme are:

- Basic requirements are those expected requirements or attributes (or "musts") whose absence results in extreme customer dissatisfaction. However, enhancing this kind of attributes does not contribute to customer satisfaction.
- Satisfiers, on the other hand, will improve customer satisfaction. Their absence or weak performance decreases customer satisfaction.
- Delighters or excitement attributes are unspoken and unexpected but, if present, they significantly increase customer satisfaction. Their absence is indifferent to customers.
- Indifferent requirements or attributes have no effect on customer satisfaction.

The general process for the definition of the product portfolio consists in [2]:

1. Identify customer requirements.
2. Construct a questionnaire.
3. Perform a survey.
4. Analyze and interpret the data.

However, the Kano Model does not include specific steps or procedures to perform the survey, and analyze and interpret the results. The purpose of our work was to complement the process defined above. By means of a case study a series of steps for performing the survey and data analysis were defined and applied. In this way, a complete process for the definition of the product portfolio using the Kano Model was obtained. Both, the process and the case study, are described in the following section.

## 3. Product definition process using the Kano Model

This section describes the process for the definition of the PL portfolio using the Kano Model. As was explained in the introduction, Statistical Consulting course students took part in this work[†]. Before explaining the process, some background about the case study is given.

### 3.1 Case study

The selected case study for the process definition was a PL for Ancora-Soft. Ancora-Soft is a case tool that supports ANCORA [6]. ANCORA is a requirements analysis methodology which has been widely applied. ANCORA has been used in requirements analysis for systems between 100 and 600 function points. ANCORA provides with several tools that allow adequate stakeholders' involvement in the process. The main modeling tool of ANCORA is the script. Like in a theater play, the cast (users) will play several roles. With the requirements analyst's help, the script is written by the stakeholders. Other main feature is that the methodology promotes reuse. ANCORA provides with several ways of storing elements from previous systems so that they can be used in the development of new systems.

### 3.2 Process

The specific activities defined for this process were the following:

1. Goals definition. Both general and specific goals were defined. In order to

[†] Aguilar, E., Bariguete, J. L. Cardenas, M. G., García, D., Landa, T., M. J, Lopez, V. I., Vazquez, O. F. J., Yoval, F. L. R

this, the expected results were clarified. In this case, it was important to determine if among the SPL desired characteristics, the SPL should also be evolvable. From SPL-specific quality attributes [7] evolvability was chosen.

2. Method and measurement instrument selection. Considering the goals already defined, several methods and measurements instruments were analyzed. Considering Kano Model requirements, a questionnaire was selected.
3. Questionnaire elaboration. The Kano Model indicates the elaboration of functional and dysfunctional questions [2] [3]. To this, a set of 32 requirements was obtained. These requirements were obtained from sources such as [6] and related publications [8] [9]. A review of available CASE tools was performed also. In this phase requirements correspond actually to high level requirements or features, and there is no dependency among the 32 requirements obtained. However, when these requirements are further analyzed and specified during the following phases, they will generate more than 32 requirements. According to the Kano Model, for each requirement a functional and a dysfunctional question were formulated. The answers to both questions allow the classification of the corresponding requirements according to Kano scheme. Since functional and dysfunctional questions can be confusing, special care was given to the elaboration of each one of them. Scientifically, there is no limit to the number of questions that can be elaborated for a questionnaire.
4. Key question identification (based on the goals definition). In the questionnaire, there is always a question that captures almost everything that is desired to know from the respondents. The customer or researcher must determine which is the most important. He/she frequently answers that all of the questions are important. In this sense it must asked that: if you only had one minute with your respondent, what will you ask? So, the key question is easily determined.
5. Determination of the target population. The target population is the population about which the inference is going to be done. This population can be finite or infinite. In this case the population corresponds to software developers who know and use ANCORA. The total of the target population is 158.
6. Sample frame determination. A list of study units has to be obtained. In this case the units of study correspond mainly to software developers, requirements analysts and project leaders.
7. Exclusion criterion definition. This criterion serves to exclude from the survey, respondents that could provide useless answers to the study. In this case the exclusion criterion was that the respondents knew ANCORA.
8. Survey data base design. The database contains all the variables that are measured in the questionnaire. This database contains 92 variables, from which 64 correspond to the functional and dysfunctional questions. This is because 32 ANCORA requirements were considered in the survey. The database is implemented in Microsoft Excel.
9. Pilot sample size calculation. This was determined as a convenience sample. This is a form of non-probability sample in which the researcher selects readily available respondents. In this study, 8 available and close software developers were surveyed about the key question. In this way, the final sample size was calculated [10].
10. Pilot survey.  Thus survey was performed to obtain information about two important aspects of the survey:
    a. Questionnaire adequacy. The need for some changes was identified. These changes were related to the respondents' general data.
    b. Variation estimation. This estimation is used to calculate the final sample size. Here, $\hat{p} = 0.42$ for the success answer of the key question [10].
11. Selection of a sampling design method. This is done taking into consideration steps 1, 3, 4, 5, and 6 above. According to the target population's distribution, the method selected was the stratified sampling method with allocation proportional to the size of each one of the software developer groups [10]: Informatics students, Software

Engineering Diploma graduates, and Master in Software Engineering students.

12. Sample size calculation. The formula to be used is determined by the key question, the kind of population and the results obtained in 10.b. The sample size can be obtained using the maximum variance ($p = 0.5$), or using the key question's estimated variation ($\hat{p} = 0.42$). Both methods were employed in order to make a comparison. Several levels of confidence and precision were used for both methods. The formula used was obtained with a finitude correction factor [10]. Such formula is:

$$n = \frac{N\hat{p}\hat{q}}{(N-1)\frac{B^2}{Z^2_{\alpha/2}} + \hat{p}\hat{q}} = 43$$

Where:

$\hat{p}\hat{q}$ = Is the key question's estimated variation or in other case the maximum variation $pq$ .

$Z$ = Confidence of 95% (1.96).

$B$ = Precision of 0.13.

$N$ = Total of available software developers 158.

$n$ = Sample size to survey 43.

**Table 3.1. Population's distribution and sample.**

| Group | Description | Total: $N_i$ | Sample $n_i = \left(\frac{N_i}{N}\right) \times n$ |
|---|---|---|---|
| 1 | Informatics students | 113 | 30 |
| 2 | Software Engineering Diploma graduates | 34 | 9 |
| 3 | Master in Software Engineering students | 11 | 4 |
| | Total | 158 | 43 |

When there is no pilot sample, the sample size can be obtained assuming maximum variance [10]. In this case, both methods obtained similar results. Given the target population's distribution, the proportional allocation was as indicated in Table 3.1.

The selection of software developers to survey was done using systematic sampling for each group and using the exclusion criterion.

13. Data analysis. Data were recorded in Microsoft Excel and new variables related to the SPL requirements were created. These new variables helped to re-code the combined questions of the Kano Model. See Table 3.2. From the new variables containing the labels of Table 3.2, the corresponding frequencies and percentages were calculated. Afterwards, a correspondence analysis was performed [11] [12]. This analysis evaluates the association between the opinions about SPL requirements and the software developers' general characteristics.

**Table 3.2. Table for data classification.**

| Functional | Dysfunctional | | | | |
|---|---|---|---|---|---|
| | Like | Expected | Don't care | Can live with it | Dislike |
| Like | ¿? | 3 | 3 | 3 | 2 |
| Expected | 5 | 4 | 4 | 4 | 1 |
| Don't care | 5 | 4 | 4 | 4 | 1 |
| Can live with it | 5 | 4 | 4 | 4 | 1 |
| Dislike | 5 | 5 | 5 | 5 | ¿? |

Where:

1. Basic requirements
2. Satisfiers
3. Delighters
4. Indifferent requirements
5. Undesired requirements

14. Results. The distribution of the respondent's ratings is analyzed for each requirement. The category is determined according to the highest rating obtained by the requirement. When categories are tied, the model employs the following rule 1>2>3>4. The resulting distribution is indicated in Table 3.3

If Kano's rule was applied there will be no variability in the requirements classification, most of the requirements will fall in the 3 category (Satisfier). Here is where the model shows its lack of statistical foundation. Therefore a Correspondence analysis was used via STATISTICA [13].

**Table 3.3. Requirements classification distribution.**

| **Require-ments** | **Classification** | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **1** | **2** | **3** | **4** | **5** |
| **R1** | 10.00 | 4.00 | 42.00 | 28.00 | 10.00 | 6.00 |
| **R2** | 0 | 12.24 | 36.73 | 30.61 | 16.33 | 4.08 |
| **R3** | 4.00 | 6.00 | 28.00 | 12.00 | 36.00 | 14.00 |
| **R4** | 2.04 | 8.16 | 28.57 | 22.45 | 20.41 | 18.37 |
| **R5** | 2.04 | 8.16 | 57.14 | 12.24 | 16.33 | 4.08 |
| **R6** | 4.00 | 16.00 | 26.00 | 14.00 | 22.00 | 18.00 |
| **R7** | 4.00 | 8.00 | 24.00 | 20.00 | 30.00 | 14.00 |
| **R8** | 2.00 | 12.00 | 46.00 | 22.00 | 16.00 | 2.00 |
| **R9** | 4.08 | 12.24 | 36.73 | 16.33 | 16.33 | 14.29 |
| **R10** | 6.12 | 14.29 | 42.86 | 18.37 | 12.24 | 6.12 |
| **R11** | 2.08 | 16.67 | 45.83 | 20.83 | 14.58 | 0.00 |
| **R12** | 4.00 | 18.00 | 40.00 | 22.00 | 14.00 | 2.00 |
| **R13** | 14.29 | 10.20 | 20.41 | 16.33 | 32.65 | 6.12 |
| **R14** | 8.00 | 8.00 | 20.00 | 20.00 | 26.00 | 18.00 |
| **R15** | 6.00 | 12.00 | 20.00 | 20.00 | 26.00 | 16.00 |
| **R16** | 4.00 | 12.00 | 24.00 | 20.00 | 26.00 | 14.00 |
| **R17** | 4.08 | 10.20 | 42.86 | 12.24 | 12.24 | 18.37 |
| **R18** | 6.00 | 8.00 | 34.00 | 6.00 | 26.00 | 20.00 |
| **R19** | 14.29 | 10.20 | 20.41 | 14.29 | 36.73 | 4.08 |
| **R20** | 12.24 | 2.04 | 59.18 | 8.16 | 8.16 | 10.20 |
| **R21** | 10.00 | 8.00 | 30.00 | 28.00 | 12.00 | 12.00 |
| **R22** | 14.00 | 12.00 | 26.00 | 20.00 | 24.00 | 4.00 |
| **R23** | 10.00 | 8.00 | 46.00 | 10.00 | 16.00 | 10.00 |
| **R24** | 14.00 | 8.00 | 46.00 | 18.00 | 10.00 | 4.00 |
| **R25** | 10.00 | 12.00 | 36.00 | 26.00 | 8.00 | 8.00 |
| **R26** | 10.00 | 12.00 | 28.00 | 26.00 | 20.00 | 4.00 |
| **R27** | 4.00 | 14.00 | 34.00 | 26.00 | 6.00 | 16.00 |
| **R28** | 14.00 | 4.00 | 26.00 | 22.00 | 22.00 | 12.00 |
| **R29** | 0 | 10.00 | 44.00 | 20.00 | 16.00 | 10.00 |
| **R30** | 8.00 | 10.00 | 44.00 | 14.00 | 18.00 | 6.00 |
| **R31** | 4.00 | 16.00 | 30.00 | 14.00 | 16.00 | 20.00 |
| **R32** | 4.00 | 14.00 | 58.00 | 10.00 | 14.00 | 0.00 |

Correspondence analysis is a descriptive technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to those produced by Factor Analysis techniques, and they allow exploring the structure of categorical variables included in the table Greenacre [14]. These methods were originally developed primarily in France by Jean-Paul Benzécri in the early 1960's and 1970's.

The purpose of correspondence analysis is to reproduce the distances between the row and/or column points in a two-way table in a lower-dimensional display. As in factor analysis, the actual rotational orientation of the axes is arbitrarily chosen so that successive dimensions "explain" less and less of the overall *Chi*-square value (or inertia). The best way to quickly review the row and column coordinates (Table 3.3) is to scatterplot them. A scatterplot (see Figure 3.1) shows the row and column points in the two dimensions [15]. In this way, one can make meaningful interpretations about the general locations of row and column points, and their relations within each type of point. In Figure 3.1, the following groups of requirements can be identified:

- Satisfiers: R1, R2, R5, R8, R10, R11, R12, R17, R20, R24, R25, R29, R30 and R32.
- Delighters: R21, R22, R26, R28.
- Indifferent: R7, R13, R14, R15, R16 and R19.
- Undesired: R4, R6 and R18.

Requirements R9, R23, R27 and R31 show no association; therefore they can not be categorized. Here, some other mechanism for classification should be used. Groups of requirements do not depend among them. The classification, however, did depend on the general characteristics of the respondents.

In this case, the characteristics that influenced the classification were: gender, age, profession, how he/she learned ANCORA, and the period of time of knowing ANCORA (*Chi*-square, $p < 0.001$). Only for R2 and R28 contradictory answers (column C) were not found, which is expected in this kind of surveys. Besides, there is not enough statistical difference in its percentage distribution. It is important to note that no basic requirements were found. One of the reasons for this could be that since currently there is not automatic support available for ANCORA, potential users find the possibility of a CASE tool really satisfying.

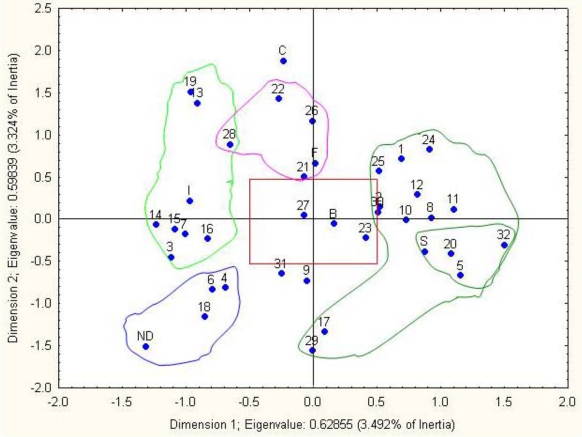

**Figure 3.1 Scatterplot of requirements distribution.**

This is a great opportunity for having an impact among the potential users. However, when looking with detail at the requirements that fall in the Satisfiers category, R5, R20 and R32 have more that 50% in the satisfier's category, almost 60%. The remaining requirements fall between 36% and 46%. This inertia allows classifying these latter requirements as basic, and R5, R20 and R32 remain as satisfiers.

With this analysis, and considering the potential groups of users (students, requirements analysts, project leaders and administrators) the product portfolio defined for Ancora-Soft is:

Product 1: All basic requirements plus satisfier R20 (to provide a high customer satisfaction)

Product 2: All basic and satisfier requirements.

Product 3: Product 2 plus all delighter requirements.

## 4. Conclusions

The Kano Model has proven to be useful in determining products' characteristics. However it lacks of statistical elements and principles to perform the overall process. Therefore, it has been complemented in the realization of the survey and in the interpretation of the data by using statistical tools and methods. The objective of this work was to integrate such elements along with the Kano Model in a process for SPL portfolio definition.

However one drawback found is that the use of functional and dysfunctional questions seems to be confusing to respondents. In this sense, is convenient to consider a validation method for the questionnaire. Other drawback found was the rule the Kano Model uses for assigning categories to the distribution. Here a scatterplot of correspondence analysis was used.

In the academic context, the involvement of Statistical Consulting students has been very enriching. The students have been able to experience the application of statistical knowledge in a real case. This case shows the use of statistical thinking for business. In this same context, a software development project has been started in order to implement a tool supporting this process.

For the development of Ancora-Soft, this study has been interesting. The use of the Kano Model has provided a new way knowing stakeholders' needs.

## 5. References

[1] P. Clements, and L. Northrop, *Software Product Lines: Practices and Patterns*, Addison-Wesley, USA, 2001.

[2] K. Pohl, G. Böckle, and F. van der Linden, *Software Product Line Engineering, Foundations, Principles, and Techniques*, Springer-Verlag, Berlin Heidelberg, Germany, 2005.

[3] N. Kano, "Attractive quality and must-be quality", *The Journal of the Japanese Society for Quality Control*, April 1984, pp. 39-48.

[4] R. W. Hoerl, "Introductory Statistical Education: Radical Redesign is Needed, or is it?", *Newsletter for the Section on Statistical Education of the American Statistical Association*, Vol. 3, No. 1, 1997.
Online at:
www.amstat.org/sections/educ/newsletter/v3n1/Hoerl.html

[5] L. López L., "La educación superior en el Siglo XXI requiere pensamiento estadístico", *Revista Cubana de Educación Superior*, La Habana, Cuba, 2007. In printing.

[6] A. Sumano Lopez, "Método para el análisis de requerimientos de software con un enfoque psicológico, social y lingüístico conducente al reuso", PhD Dissertation, IPN, Mexico, 2002.

[7] Anastasopoulos M., J. Bayer J., "Product Family Specific Quality Attributes", Fraunhofer Institut Experimentelles Software Engineering, IESE Report No. 042.02/E, Version 1.0, Kaiserslautern, Germany, 2002.

[8] E. Ortega Carreon, "ANCORASOFT: Herramienta CASE basada en componentes para el análisis de Requerimientos de Software", B.Sc. Dissertation, UV, México, 2007.

[9] E. Ortega Carreon, and A. Sumano Lopez, "Ancora_Soft: Herramienta CASE basada en componentes para el análisis de Requerimientos de Software", Primer Coloquio en Tecnologías de Información, Universidad Veracruzana, Xalapa, Ver. Mexico, 2007.

[10] S. L. Lohr, *Muestreo: Diseño y Análisis*, Internacional Thomson Editores, México, 2000.

[11] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Análisis Multivariante*, Prentice Hall, New York, 1999.

[12] R. A. Jhonson, and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, New York, 1998.

[13]Stat Soft. Inc (2004). STATISTICA (data analysis software system), version 7. Online at: www.statsoft.com.

[14] M. J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, London, UK, 1984.

[15] M. J. Greenacre, *Correspondence Analysis in Practice*. Academic Press, London, UK, 1993.