Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

A statistical model for shelf life estimation using sensory evaluation scores

Marta A. Freitas, Wagner Borges and Linda Lee Ho

Relatório Técnico
RTP-10/2001

Relatório Técnico
Série Pesquisa

A statistical model for shelf life estimation using sensory evaluations scores.

Marta Afonso Freitas (marta@est.ufmg.br)
Departamento de Estatística, ICEx, UFMG, Brazil

Wagner Borges (wborges@ime.usp.br)
Departamento de Estatística, IME, USP, Brazil

Linda Lee Ho (linda@ime.usp.br)
Departamento de Engenharia de Produção, EP, USP, Brazil

Abstract

This article focuses on the problem of estimating the shelf life of food products by modeling the results coming from sensory evaluations. In such studies, trainned panelists are asked to judge food attributes by reference to a scale of numbers (scores varying often from 0 to 6). The usual statistical approach for data analysis is to fit a regression line relating the scores and the time of evaluation. The estimate of the shelf life is obtained by solving the regression equation and replacing the score by a cut-off point (which indicates product "failure") previously chosen by the food company. The procedure used in these sensory evaluations is such that one never knows the exact "time to failure". Consequently, data arising from these studies are either right or left censored. We propose a model which incorporates these informations and assumes a Weibull for the underlying distribution of the failure time. Simulation studies were implemented. The approach was used in a real data set coming from sensory evaluations of a dehidrated food product.

KEY WORDS: *left censored; right censored; sensory evaluations; shelf life; Weibull distribution.*

1 Introduction

The quality of food products is a fragile thing, because of their own nature. They are susceptible to spoilage, loss of nutrients, and changes of sensory properties such as color, flavor and odor. Consequently stability studies are routinely conducted in the food industry as a part of each product development program, whether it includes new product, a product improvement, or a change in type or specification of an ingredient. Some studies center on the rate of degradation, and others on estimating the *shelf life*: the length of time required for the product to be unfit for human consumption. By *unfit for human consumption* it is meant that the product exhibits either physical, chemical, microbiological, or sensory characteristics that are unaccepted for regular consumption. The manufacturer attempts to develop a product with the longest shelf life practical, consistent with costs and the pattern of handling and use by distributors, retailers, and consumers. Inadequate shelf life determination will lead to consumer dissatisfaction or complaints. At best, such dissatisfaction will eventually affect the acceptance and sales of brand name products. At worst, it can lead to malnutrition and even illness. For these reasons, food processors pay great attention to adequate storage stability or shelf life.

A very practical use of product shelf-life information is for open-dating purposes. Open dates are placed on the labels of foods and beverage products to assist consumers in the overall process of decision-making and purchase. A variety of dates are currently used; with the exception of pack date, each is based on the shelf life stablished for the product. Each country has its own regulation, but the most commum open dates are the following:

- **sell-by or pull date**: the date after which the food can no longer be sold as fresh; the last date a retail store should offer the food for sale provided it has been stored and handled properly; but not the last day it can be eaten without loss of quality. To allow for a reasonable period of home storage and use, the pull date should be considerably earlier than the end of shelf life;

- **use-by ou expiration date**: the last day a food may be acceptable for its intended use;

- **freshness or quality assurance date**: the date after which the product falls below optimum quality level. After this date, the food will be edible but will lose quality. Good manufacturing practice would set this date earlier than the end of shelf life to allow for normal variations in home storage conditions;

- **pack date**: the date the product was packed or packaged in the immediate container in which it will be sold. It does not provide any specific information as to the quality of the product when purchased or how long it will retain its quality after purchase.

When one talks about determining the shelf life, chemical, physical, microbiological, and nutritional analyses are fundamental but equally important are the sensory characteristics of the product. For this reason, sensory evaluations are conducted in food experimentations to determine the shelf life. In such experiments, a sample of product units is stored and periodically, at preespecified evaluation times, a sample of units is collected from the ones stored and subjected to sensory evaluations by a trained panel. Each panelist is asked to judge each product's atribute separately by reference to a rating scale, for instance, a seven point rating scale varying from 0 to 6. Because of the destructive nature of the evaluations, units that have already been evaluated at a given time cannot be restored to be evaluated later on. The usual statistical approach to analyse this kind of data and estimate the shelf life is to fit a regression line relating the scores (y) and the preespecified times of evaluation (x). An estimate of the shelf life is obtained by solving the fitted regression equation for x and replacing the score (y) by a cut-off point (which indicates product "failure") previously chosen by the food company. Gacula and Singh (1989) present examples in which regression models are implemented to estimate the shelf life. However, this approach offers dificulties since in general, the assumption of normality and homocedasticity - basic requirements of regression analysis (Draper and Smith, 1998) - are not valid for sensory scores (Gacula and Singh, 1989). Gacula and Singh (1984) suggest some transformations of the experimental data to a new scale where the normality assumption may be approximately satisfied. Alternatives to overcome the violation of the constancy of variance assumption include the use of variance-stabilizing transformations and weighted least squares (Draper and Smith, 1998). However, it seems that these two procedures are not very helpfull to overcome the problem with this particular kind of data (Gacula and Singh, 1984).

An alternative approach is to fit a parametric lifetime model such as Weibull, lognormal, to the failure data (Meeker and Escobar, 1998). But, in these experiments, one does not have the "time to failure" of a given unit. For a unit evaluated at a preestablished time, one of the two situations can happen: the score is either less than or equal to the cut-off point or greater than the cut-off point chosen by the food company. In the former case the "failure time" is somewhere between the start of the experiment and the present time of evaluation. In the latter case, the product is still good for consumption and its "failure" will take place sometime after evaluation. Thus, data coming from these sensory evaluations are either left or right censored (Lawless, 1982). Gacula (1975) and Gacula and Kubala (1975) suggest the use of a "staggered design" to implement weekly sensory evaluations. The intervals between evaluations are larger at the beginning of the experiment and smaller towards the end. The number of units sampled increases towards the end of the experiment. The idea is to get closer to the "real" failure time. Gacula and Kubala (1975) present examples of data coming from these experiments. Evaluation times are used as approximations to the real failure time. Shelf life is then estimated using the Weibull distribution. Two problems can be pointed out: (1) the failure time model (in this case, Weibull) is not used appropriately since the evaluation times are fixed and (as a consequence) (2) there is a great risk of underestimating the shelf life.

The present investigation was motivated by a real situation experienced by the authors and, to overcome those difficulties, an alternative approach is proposed. In this approach the right and left censored data are incorporated naturaly. In addition, the model assumes a Weibull for the underlying distribution of the failure time. By using this model, all the information available is incorporated.

The paper is organized as follows. In Section 2 we introduce the motivating (real) situation along with some background information. In Section 3, the proposed model is presented. The description of the simulation study and the results itself are in Section 4. In Section 5 the proposed procedure is applied to analyse the sensory data coming from the motivating situation. The paper concludes with a discussion.

2 Motivating situation

Sensory evaluations were conducted by a food company at the laboratory level in order to determine the shelf life of a manufactured dehidrated product, stored at different environmental conditions. Three attributes were evaluated by trained panelists: odor, flavor, appearance. The main characteristics of this study are presented next.

Experimental design. A lot of product units was sampled from the production line and the units were randomly assigned to one of the following storage conditions:

- refrigeration: these units were kept under refrigeration at 4^oC (approximately). Temperature and humidity levels were not controlled but average weekly values were recorded continuously. These units were used as reference (control) during the trials;
- room temperature and humidity: average weekly levels were recorded continuously;
- environmental chamber 1: temperature and humidity levels controlled at 30^oC and 80% repectively;
- enviromental chamber 2: temperature controlled at 37^oC. Humidity levels were not controlled but average weekly values were recorded;

The last two conditions were used in order to simulate an aggressive storage environment. Researchers expected to register a shorter shelf life under those conditions when compared with storage under room temperature. In addition to the variables mentioned (relative humidity, temperature), the **water activity** in each environment (a_w) was also measured. This parameter is defined by the ratio of the water vapor pressure of food substrate to the vapor pressure of pure water at the same temperature $a_w = p/p_0$ where p=vapor pressure of solution and p_0 = vapor pressure of solvent (usually water). It is now generally accepted that the water requirements of microorganisms should be defined in terms of the water activity (a_w). The a_w of most fresh foods is above 0.99. In general, bacteria require higher values of a_w for growth than fungi, for example. Most spoilage bacteria do not grow below a_w=0.91, while spoilage molds can grow as low as 0.80. Jay (1992) presents a table with approximate minimum a_w values for growth of important microorganisms in foods. Some relationships have been shown to exist among a_w values, temperature and nutrition. That is why those values where also recorded during the experiment.

Laboratory Panel. Forty five (45) subjects were trained for the sensory characteristics of the product before the main trial started. Sensory evaluations were made initially and then every week thereafter. Each week, near eight trained subjects were selected to form the test panel.

Test Procedure. Evaluations were performed weekly. Each week, one unit was sampled from each one of the four storage conditions. Each panelist was offered simultaneously a set of three units: one labeled as reference (control) and other two test units labeled with a three digits number. One of the two test units was always a "blind" reference. Therefore, at a given week, each panelist was offered in random order, three sets of units to be evaluated, namely [RE, BRE, R]; [RE, BRE, CH1] and [RE, BRE, CH2], where RE, BRE and R stand respectively for "reference","blind reference" and "room temperature and humidity"; CHi stands for "chamber i" i=1,2. Within a given group, the reference unit (RE) was always evaluated first. For the other two (BRE, R, CH1 or CH2), the order was randomized. All units were descarded after evaluation.

Measurement Scale. Panelists were asked to compare each test unit (including the "blind" reference) with the reference and assign a score on a seven-point scale (0 to 6) individually to each attribute: 6 = "no difference"; 5 = "very slight difference"; 4 = "slight difference"; 3 = "different"; 2=" large difference" 1= "very large difference"; 0="total difference".

Criterion of Failure. The manufacturer adopted the following failure criterion: for each attribute, product units scored 0,1,2 or 3 were considered unfit for human consumption.

Follow up time: units stored at room temperature; chambers 1 and 2 were followed for 51, 36 and 18 weeks respectively. Table 1 shows panel scores for units stored in chamber 1 (30^oC; 80%), evaluated at the 13^{th} week.

Table 1: Panel scores for units stored in environmental chamber 1 (30^oC;80%)

13^{th} week		panelist (code number)						
attribute	storage condition	28	21	26	25	1	13	2
odor	refrigerated**	6*	6	6	6	6	6	6
	CHAMBER 1	6	6	6	3	3	5	6
flavor	refrigerated	6	5	6	6	6	6	6
	CHAMBER 1	6	5	6	3	3	4	2
appearance	refrigerated	6	6	6	6	6	6	6
	CHAMBER 1	6	6	6	5	5	6	4

(*) Scale: from 0 ("total difference")to 6 ("no difference"); (**) "blind" reference.

It should be noticed that the blind references were used only to check the consistency of painelists' judgement. In other words, there was no interest in studying that storage condition.

Since attributes were scored separately, it is possible to have a product unit classified as *unfit* regarding one particular attribute and *fit* regarding another one.

According to the failure criterion adopted by the manufacturer, at a given preestablished evaluation week one of the following situations (for each atribute)might happen: if the attribute's score is less than or equal 3 (three) then one knows that the particular unit has become *unfit* for human consumption in a moment somewhere between the beginning of the trial and the evaluation week. On the other hand, if the attribute's score is greater than 3 then, that unit is still *fit* for consumption (regarding that attribute). Unfortunately, because of the trial's destructive characteristic, the follow up of that unit is interrupted.

To estimate appropriately the shelf life of that food product it is desirable to incorporate both the information contained in the right and left censored data. We will address this problem in the following sections.

3 A Statistical Model for Sensory Data

Suppose a sample of $N = \sum_{i=1}^{k} n_i$ food product units is taken from the production line and stored under a given environmental condition. These units will be evaluated by a trained panel at prestablished evaluation times in order to determine its shelf life.

Let τ_i $(i = 1, ..., k)$ be the evaluation times (fixed). Then, at the evaluation time τ_1, n_1 units are sampled from the total N and subjected to a sensory evaluation by n_1 panelists who score each attribute (odor, flavor, appearance) using a 7 point rating scale (for instance, 0 to 6).

The evaluation is destructive, consequently these n_1 units can no longer be followed in time. Next at τ_2, n_2 units are sampled from the $N - n_1$ units left and evaluated by n_2 painelists. This process is repeated through the last evaluation time τ_k when the remaining n_k units are finally evaluated.

Let Z_{ij} be the score assigned to the j^{th} product unit ($j = 1, ..., n_i$) evaluated at the time τ_i ($i = 1, ..., k$).

Each of the n_i units evaluated at a given time will be considered as *unfit for consumption* (regarding a particular attribute) depending on the score assigned by the panelist. Let us refer for a moment to the real situation described in Section 2. In that case, the j^{th} unit evaluated at time τ_i will be considered unfit for consumption (regarding the attribute being evaluated) if Z_{ij}=0,1,2 or 3. If, on the other hand, $Z_{ij} > 3$ then the attribute "failure" will occur sometime in the future but we will not be able to know when at the present time. It should be enfasized that it is possible to have a unit being considered "unfit" regarding its flavor but "fit for consumption" regarding its appearance.

We can define a new random variable Y_{ij}, given by

$$Y_{ij} = \begin{cases} 1 & \text{if } Z_{ij} < 4 \text{ (i.e., if score} < 4) \\ 0 & \text{if } Z_{ij} \geq 4 \text{ (i.e., if score} \geq 4). \end{cases}$$

Therefore, at each fixed time τ_i we have a random sample of size n_i from a random variable Y_{ij} where Y_{ij} is Bernoulli distributed with probabiity p_i given by

$$p_i = P(Y_{ij} = 1) = P(0 < T_{ij} \leq \tau_i)$$

,
where T_{ij} is the failure time of the j^{th} unit evaluated at τ_i.

Thus, in a equivalent way, Y_{ij} can be defined as

$$Y_{ij} = \begin{cases} 1 & \text{if } 0 < T_{ij} \leq \tau_i \\ 0 & \text{if } T_{ij} > \tau_i. \end{cases}$$

Or

$$\begin{aligned} P(Y_{ij} = s_{ij}) &= P(0 < T_{ij} \leq \tau_i) = 1 - R(\tau_i), \text{ if } s_{ij} = 1 \\ &= P(T_{ij} > \tau_i) = R(\tau_i), \text{ if } s_{ij} = 0, \end{aligned} \tag{1}$$

where R(.) in (1) is the **reliability function** (Nelson, 1990).

The main purpose here is to estimate the shelf life of a food product, taking into account some sensory quality characteristics. If we take a better look at this problem, we see that in fact the shelf life itself is a random variable whose behavior for each attribute follows some underlying distribution. One way to tackle this problema is to estimate percentiles of each shelf life distribution (considering each attribute separately) and then pick one for each attribute to represent the "attributes' shelf life". If the manufacturer decides to have only one value reported as a shelf life for the product, the minium value could be chosen.

What is really important is to estimate characteristics of the shelf life distribution, such as percentiles, fraction "defective ou unfit" at a given time t_0 (given by $F(t_0) = 1 - R(t_0)$), to list a few. That can be done by postulating an underlying distribution for the failure time T and writing down a likelihood function taking into account the distribution of the random variable Y_{ij}. This approach is presented in Section 3.1.

3.1 Weibull failure time distribution

It is assumed that the failure time T_{ij} of the j^{th} unit evaluated at time τ_i (fixed) has a Weibull distribution, with parameters α_j e $\delta \geq 1$, and

- the parameters α_j and δ are defined by

$$\begin{aligned} \alpha_j &= \exp\{\boldsymbol{X_j\beta}\} = \exp\{X_j^0\ \beta_0 + X_j^1\ \beta_1 +\ \ldots\ + X_j^q\ \beta_q\}, \\ (\ j &= 1,2,...,n_i\ ) \quad \text{and} \quad \delta = \exp(\gamma) \quad \gamma \geq 0 \end{aligned} \qquad (2)$$

- $\boldsymbol{X_j} = (X_j^0, X_j^1, ..., X_j^q)$ is a (q+1) vector of covariates related to the j^{th} unit evaluated at τ_i
- $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_q)'$ is a (q+1) vector of parameters associated to the covariates.

Now using the fact that for a Weibull distribution with parameters α and δ, $R(t) = \exp\{-(\alpha\ t)^\delta\}$ and Y_{ij} has a Bernoulli distribution with probability p_i given by (1), the likelihood function is given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left\{ \left[e^{-(\tau_i\ \alpha_j)^\delta}\right]^{1-s_{ij}} \quad \left[1-e^{-(\tau_i\ \alpha_j)^\delta}\right]^{s_{ij}} \right\} \\ &= \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left\{ \left[e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right]^{1-s_{ij}} \quad \left[1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right]^{s_{ij}} \right\}, \end{aligned} \qquad (3)$$

where $\boldsymbol{\theta'} = (\boldsymbol{\beta'}; \delta)$.

Maximum likelihood estimates are obtained by direct maximization of the logaritmum of expression (3). The expressions of the first derivatives are:

$$\left[\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{(q+2)\times 1} = \begin{bmatrix} \frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} \\ \frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\gamma} \end{bmatrix}, \qquad (4)$$

where:

$$\begin{aligned} \frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\boldsymbol{X_j'}\left\{-(1-s_{ij})e^\gamma(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma} + \frac{s_{ij}e^\gamma e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}{\left[1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right]}\right\}, \\ \frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\gamma} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{e^\gamma(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}\ln(\tau_i e^{\boldsymbol{X_j\beta}})\left[-(1-s_{ij}) + \frac{s_{ij}e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}}{\left(1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right)}\right]\right\}. \end{aligned}$$

The elements of the Fisher Information matrix I$(\boldsymbol{\theta})$ $((q+2)\times(q+2))$ are given by

$$I(\boldsymbol{\theta}) = E\left\{-\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]\right\}_{(q+2)\times(q+2)} = \begin{bmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{bmatrix}, \tag{5}$$

where:

$$I_{11} = E\left\{-\left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]\right\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{X_j'X_j})\left\{\frac{e^{2\gamma}e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}(\tau_i e^{\boldsymbol{X_j\beta}})^{2e^\gamma}}{\left(1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right)}\right\},$$

$$I_{12} = E\left\{-\left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \gamma}\right]\right\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\boldsymbol{X_j'}\left\{\frac{e^{2\gamma}e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}(\tau_i e^{\boldsymbol{X_j\beta}})^{2e^\gamma}\ln(\tau_i e^{\boldsymbol{X_j\beta}})}{\left(1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right)}\right\},$$

$$I_{22} = E\left\{-\left[\frac{\partial^2 \mathcal{L}}{\partial \gamma^2}\right]\right\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\{\frac{e^{2\gamma}e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}(\tau_i e^{\boldsymbol{X_j\beta}})^{2e^\gamma}[\ \ln(\tau_i e^{\boldsymbol{X_j\beta}})\ ]^{\ 2}}{\left(1-e^{-(\tau_i e^{\boldsymbol{X_j\beta}})^{e^\gamma}}\right)}\right\},$$

with I_{11}, I_{12} and I_{22} with dimensions $(q+1)\times(q+1)$, $(q+1)\times 1$ and 1×1 respectively.

Maximum likelihood estimator is obtained by implementing numeric optimization methods such as the well known Newton-Raphson algorithm. In this work we have used a minor adjustment to the Newton-Raphson procedure, sometimes used in statistical problems, called Fisher's Score (McCullagh and Nelder, 1989).

If $\hat{\boldsymbol{\theta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_q, \hat{\gamma})'$ is the maximum likelihood estimator of $\boldsymbol{\theta} = (\beta_0, \beta_1, ..., \beta_q, \gamma)'$ then

- for a given set of covariates $\boldsymbol{X_j} = (X_j^0, X_j^1, ..., X_j^q)$, the maximum likelihood estimate of the percentile $t_{p(j)}$ is given by

$$\hat{t}_{p(j)} = \frac{1}{\hat{\alpha}_j}\left[-\ln(1-p)\ \right]^{\frac{1}{\hat{\delta}}}$$

  ,

  where, $\hat{\alpha_j} = \exp\{\boldsymbol{X_j}\widehat{\boldsymbol{\beta}}\}$ and $\hat{\delta} = \exp\{\hat{\gamma}\}$

- using maximum likelihood large sample theory (asymptotic normality) and the delta method(Cox and Hinkley, 1974), it is possible to find the expression of the asymptotic variance estimator

$$\widehat{Var}(\hat{t}_{p(j)}) \doteq Z'\ I^{-1}(\boldsymbol{\theta})\ Z\ _{\left\{\text{para } \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}\right\}}, \tag{6}$$

  where Z is a vector of dimension (q+2) $\times$ 1 given by

$$Z = \begin{bmatrix} \left(\frac{-(-\ln(1-p)\ )^{\frac{1}{e^\gamma}}}{e^{\boldsymbol{X_j\beta}}}\right)\ \boldsymbol{X_j'} \\ -------------- \\ \frac{-[-\ln(1-p)\ ]^{\frac{1}{e^\gamma}}\ [\ e^{-\gamma}\ln(-\ln(1-p)\ )\ ]}{e^{\boldsymbol{X_j\beta}}} \end{bmatrix}. \tag{7}$$

Therefore, 95% confidence interval (asymptotic) for $t_{p(j)}$ is

$$UB = \hat{t}_{p(j)} + (1,96)\left\{\widehat{Var}(\hat{t}_{p(j)})\right\}^{1/2} , \quad (8)$$

$$LB = \hat{t}_{p(j)} - (1,96)\left\{\widehat{Var}(\hat{t}_{p(j)})\right\}^{1/2} \quad (9)$$

- equivalently, for a given set of covariates $\boldsymbol{X_j} = (X_j^0, ..., X_j^q)$, the maximum likelihood estimator of the fraction defective in t_0, $F_j(t_0)$, is given by

$$\hat{F}_j(t_0) = 1 - \hat{R}_j(t_0) = 1 - e^{-(t_0 e^{\boldsymbol{X_j}\widehat{\boldsymbol{\beta}}})^{e^{\hat{\gamma}}}} \quad (10)$$

- making use of the asymptotic normality property of the maximum likelihood estimator and the delta method (Cox e Hinkley, 1974) we get the expression of the 95% confidence interval for the fraction defective $\hat{F}_j(t_0) = 1 - \hat{R}_j(t_0)$:

$$UB = 1 - \left\{\hat{R}_j(t_0)\right\}^{\exp\left\{1,96\ [\widehat{Var}(\hat{\phi})]^{1/2}\right\}} , \quad (11)$$

$$LB = 1 - \left\{\hat{R}_j(t_0)\right\}^{\exp\left\{-1,96\ [\widehat{Var}(\hat{\phi})]^{1/2}\right\}} , \quad (12)$$

where $\hat{\phi}$ and $\widehat{Var}(\hat{\phi})$ in (11) and (12) are given by

$$\hat{\phi} = \ln(-\ln\ \hat{R}_j(t_0)\ ) , \quad (13)$$

$$\widehat{Var}(\hat{\phi}) \doteq Z'\ I^{-1}(\boldsymbol{\theta})\ Z_{\left\{\text{para } \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}\right\}} , \quad (14)$$

and Z is a vector of dimension (q+2) $\times$ 1 given by

$$Z = \begin{bmatrix} e^{\gamma}\ \boldsymbol{X'_j} \\ - - - - - - - \\ e^{\gamma}\ \ln(t_0\ e^{\boldsymbol{X_j\beta}}\ ) \end{bmatrix} . \quad (15)$$

We point out that expressions (11) and (12) were calculated applying the asymptotic normal distribution to the tranformation ϕ (expression [15])for which the range is unrestricted. Then, the confidence interval for the fraction defective is found applying the inverse transformation. This procedure suggested by Kalbfleisch and Prentice (1980; page 15) prevents the occurence of limits outside the range [0,1].

4 Simulation Study

This section presents simulation results for the Weibull model with no covariates:

$$\begin{aligned} \boldsymbol{X_j} &= (X_j^0, ..., X_j^q) = (X_j^0) \equiv (1) \quad \text{(fixed) and,} \\ \boldsymbol{\beta} &= (\beta_0, \beta_1, ..., \beta_q)' \equiv (\beta_0). \end{aligned}$$

Therefore,

$$\begin{aligned} \alpha_j &= \exp(X_j^0 \beta_0) = \exp(\beta_0) = \alpha, \\ \text{and } \delta &= \exp(\gamma) \text{ for all } j = 1, 2, ..., n_i \text{ and } i = 1, 2, ..., k. \end{aligned}$$

Consequently, $\boldsymbol{\theta} = (\beta_0, \gamma)'$ is the vector of parameters to be estimated by maximum likelihood.

4.1 Description of the Simulation Study

Some basic questions had to be answered in order to implement the study:

1) Which values of α and δ (and consequently of β_0 and γ) should be used in the simulations?

Our main purpuse was to study the "quality" or "performance" of the estimates obtained by the model proposed in situations which imitate the real data available as close as possible. Our emphasis was in getting good estimates of percentiles and fraction of defectives. As measures of "quality" or "performance", we used:

- the absolute bias B= $\{(\sum_{i=1}^{N} \hat{u}_i)/N\}$ − (real value of u) where $\hat{u}_i$ are the estimated parameter values for each of the N samples generated;
- the relative bias RB={(absolute value of B)/(real value of u)}x100%;
- the standard deviation (SD) and the mean square error (MSE).

To get a first guess of the range of values to be used in the simulation, the model proposed was fitted to the real data set. The parameters estimates are summarized in Table 2.

Table 2: Parameters estimates for the Weibull model applied to the real data

	attribute					
	odor		flavor		appearance	
storage condition	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\delta}$
room temp.and humidity	0.0191	1.3	0.0181	1.2	0.0193	1.9
environmental chamber 1	0.0302	1.6	0.0358	1.4	0.0233	2
environmental chamber 2	0.0596	1.6	0.0659	1.4	0.0602	2

For each storage condition, the average value of the estimates obtained for α ($\bar{\alpha}$) where 0.019, 0.0298 and 0.0619 for storage conditions room temperature and humidity, chamber 1 and 2 respectively. Simulations were implemented using $\alpha = 0.020$, $\alpha = 0.035$ and $\alpha = 0.065$. δ values ranged from 1.2 to 2, with incremets of size $\triangle$ =0.2. Note that the values presented in Table 2 are included in this range.

2) Which sample plans should be implemented in the simulation study?

By "sample plan" we mean: the number of weeks of follow-up (nw), the number of panelists allocated to each week (np) and the total number of product units under test ($N = nw \times np$).

Once again the idea was also to mimic the scenarios found in the real data set. Tables 3, 4 and 5 summarize the sample plans used.

Table 3: Scenarios considered in the simulation study: α =0.02 and δ = 1.2 to 2, increments of 0.2

	characteristics of each sample plan		
sample plan	$N^{(1)}$	$nw^{(2)}$	$np^{(3)}$
I	357	51	7
II	714	51	14
III	280	40	7
IV	560	40	14
V	504	36	14
VI	224	32	7
VII	448	32	14
VIII	168	24	7
IX	336	24	14

(1): number of units; (2): number of follow-up weeks;
(3): number of panelists assigned to each week.

Table 4: Scenarios considered in the simulation study: α =0.035 and δ = 1.2 to 2, incremetns of size 0.2

	characteristics of each sample plan		
sample plan number	$N^{(1)}$	$nw^{(2)}$	$np^{(3)}$
I	252	36	7
II	504	36	14
III	448	32	14
IV	392	28	14

(1): number of units; (2): number of follow-up weeks ;
(3): number of panelists assigned to each week.

Sample plans labelled "I" in Tables 3, 4 and 5 are exactly the ones implemented in the sensory evaluations of the dehidrated product mentioned in the previous sections.

Table 5: Scenarios considered in the simulation study: α =0.065 and δ = 1.2 to 2, increments of size 0.2

	characteristics of each sample plan		
sample plan number	$N^{(1)}$	$nw^{(2)}$	$np^{(3)}$
I	126	18	7
II	252	18	14
III	168	12	14

(1): number of units; (2): number of follow-up weeks;
(3): number of panelists assigned to each week.

3) Which steps will be followed in the simulation study?

In the proposed model, the underlying distribution of the failure time is a Weibull but in a real test situation, in fact what is observed is the score assigned by the panelist to a given product unit. In the model proposed in Section 3, the results are dichotomized according to the cut-off point stablished by the company. In other words, the result is either zero or one depending on the "failure" time (T) be located before ou after the evaluation time (τ).

In the simulation study we assumed that the evaluations were implemented weekly. In addition, the total follow-up time is nw weeks and np panelists are requested weekly to compose the laboratory panel.

The main steps followed are given below:

- **step 1**: choose a set of parameters values α and δ, and one of the sample plans listed in Tables 3, 4 and 5;
- **step 2**: generate a random sample of size N=nw× np of failute times t, from a Weibull (α, δ) were $\alpha = e^{\beta}$ and $\delta = e^{\gamma}$; values of α and δ were choosen in step 1;
- **step 3**: dichotomize the results and store them in a vector (Y). The dichotomization is done by comparing each of the N failure times t with the N evaluation times τ (τ assumes values from 1 to nw weeks). If t> τ then Y=0, otherwise Y=1 (failure has already occured);
- **step 4**: calculate the maximum likelihood estimator of β and γ (or, α and δ) using the dichotomized data and epression (3);
- **step 5**: using the parameters estimates calculated in step 4, find the estimates of percentiles and fraction of defectives;
- **step 6**: store the values calculated in 5;
- **step 7**: generate another random sample as in step 2 and repeat steps 3 to 6;
- **step 8**: steps 2 to 7 should be repeated until 1000 samples have been generated. Then, based on the 1000 random samples, calculate for each percentile and fraction of defective of interest:

- the average of the 1000 values estimated;
- the standard deviation (SD) based on the 1000 values;
- the absolute bias (B); the relative bias (RB) and the MSE (mean square error).

These steps were implemented for each one of the sample plans listed in Tables 3, 4 and 5.

4.2 Simulation Results

Tables 6 and e 7 present the simulation results for percentiles of a Weibull (α;δ), with $\alpha = 0.02$ and $\delta = 1.2$ and 1.6 respectively.

All the comparisons were done taking sample plan "I" as the reference, since that was the sample plan used in the real data set.

For the Weibull (0.02; 1.2) plans II, V e IV presented better results than plan I, considering all the performance measures(SD; RB and MSE).

The results obtained with plan II where expected since the sample size for this case was doubled. On the other hand, plans IV and V generated results much more precise and with smaller bias. In other words, the test could have been implemented with a shorter follow-up time (40 ou 36 weekes). But the price one would have to pay is to alocate 14 instead of 7 panelists each week.

For the Weibull(0.02; 1.6) the results are practically the same. Here, plans IV and V invert positions. Plans II, IV and V have better measures of performance (once compared to plan I) when we estimate percentiles.

Similar results where found for other values of δ, for instance $\delta = 1.4$, 1.8 and 2.

Table 8 shows the results of the simulation study considerig a Weibull with $\alpha = 0.02$ and $\delta = 1.2$. The quantities estimated were fractions of defectives calculated at different points in time.

Here, it is not possible to identify sample plans with better peformance than plan I for all measures calculated. For instance, for a Weibull (0.02; 1.2) plans II, IV and V are better than plan I when we take into account the SD and MSE (one exception is plan V for $t_0 = 32$ weeks).

For the Weibull (0.02; 1.6) plan II has better performance than plan IV and V. But, plans IV and V are not better than plan I when all measures of performance are compared. For instance, for plan IV, the bias associated with the estimate of $t_0 = 4$ and 32 weeks and the relative bias for 32 weeks are both larger than the ones obtained for plan I.

Therefore, when the quantity to be estimated is the fraction of defectives, the results have shown that for the Weibull (0.02; δ) plan II has the best performance regarding all performance measures. Plans IV and V have also good performance but it depends also on the indicator choosen for the comparison.

The results for other values of δ (1.4;1.8 and 2) are very similar to the ones just described.

Table 6: Simulation results (1000 samples)- percentiles t_p of a Weibull with $\alpha = 0.02$ and $\delta = 1.2$

plan	p	true value	estimate	SD(1)	bias(2)	RB(3) (%)	MSE
N=357	10^{-6}	0.0005	0.0015	0.0038	0.0010	201	1.5×10^{-5}
nw=51	10^{-5}	0.0034	0.0072	0.0128	0.0039	114	0.0002
np=7	10^{-4}	0.0232	0.0372	0.0459	0.0140	60	0.0023
I	10^{-3}	0.1582	0.2029	0.1698	0.0447	28	0.0308
	10^{-2}	1.0817	1.1866	0.6110	0.1049	10	0.3843
N=714	10^{-6}		0.0009	0.0012 (-68)(4)	0.0004 (-60)	86 (-57)	1.5×10^{-6}
nw=51	10^{-5}		0.0052	0.0054 (-58)	0.0018 (-54)	54 (-52)	3×10^{-5}
np=14	10^{-4}		0.0305	0.0240 (-48)	0.0073 (-48)	32 (-47)	0.0006
II	10^{-3}		0.1840	0.1033 (-39)	0.0258 (-94)	16 (-43)	0.0110
	10^{-2}		1.1531	0.4094 (-33)	0.0713 (-32)	7 (-30)	0.1738
N=280	10^{-6}		0.0020	0.0045	0.0015	295	2.2×10^{-5}
nw=40	10^{-5}		0.0089	0.0154	0.0055	161	2.7×10^{-4}
np=7	10^{-4}		0.0425	0.0545	0.0193	83	0.0033
III	10^{-3}		0.2176	0.1958	0.0594	38	0.0419
	10^{-2}		1.2142	0.6810	0.1324	12	0.4814
N=560	10^{-6}		0.0013	0.0022 (-42)	0.0008 (-20)	155 (-23)	5.2×10^{-6}
nw=40	10^{-5}		0.0066	0.0085 (-34)	0.0031 (-21)	92 (-19)	8.3×10^{-5}
np=14	10^{-4}		0.0351	0.0343 (-25)	0.0119 (-15)	51 (-15)	0.0013
IV	10^{-3}		0.1978	0.1372 (-19)	0.0397 (-11)	25 (-11)	0.0204
	10^{-2}		1.1806	0.5125 (-16)	0.0989 (-6)	9 (-10)	0.2724
N=504	10^{-6}		0.0012	0.0020 (-47)	0.0007 (-30)	142 (-29)	4.3×10^{-6}
nw=36	10^{-5}		0.0062	0.0080 (-38)	0.0028 (-28)	83 (-27)	7.2×10^{-5}
np=14	10^{-4}		0.0336	0.0331 (-28)	0.0104 (-26)	45 (-25)	0.0012
V	10^{-3}		0.1903	0.1346 (-21)	0.0321 (-28)	20 (-29)	0.0191
	10^{-2}		1.1489	0.5064 (-17)	0.0672 (-36)	6 (-40)	0.2610
N=224	10^{-6}		0.0034	0.0081	0.0029	574	7.4×10^{-5}
nw=32	10^{-5}		0.0133	0.0253	0.0099	290	7.4×10^{-4}
np=7	10^{-4}		0.0560	0.0809	0.0328	141	0.0076
VI	10^{-3}		0.2560	0.2619	0.0978	62	0.0781
	10^{-2}		1.3003	0.8265	0.2185	20	0.7308
N=448	10^{-6}		0.0018	0.0034	0.0013	255	1.3×10^{-5}
nw=32	10^{-5}		0.0082	0.0125	0.0048	142	0.0002
np=14	10^{-4}		0.0405	0.0466	0.0173	74	0.0025
VII	10^{-3}		0.2122	0.1726	0.0541	34	0.0327
	10^{-2}		1.2057	0.5993	0.1240	11	0.3745
N=168	10^{-6}		0.0082	0.0260	0.0077	1537	7.4×10^{-4}
nw=24	10^{-5}		0.0252	0.0627	0.0218	640	0.0044
np= 7	10^{-4}		0.0846	0.1563	0.0614	264	0.0282
VIII	10^{-3}		0.3170	0.4049	0.1588	100	0.1892
	10^{-2}		1.3835	1.0676	0.3018	28	1.2308
N=336	10^{-6}		0.0034	0.0087	0.0029	583	$8,5 \times 10^{-5}$
nw=24	10^{-5}		0.0132	0.0259	0.0098	287	7.6×10^{-4}
np=14	10^{-4}		0.0546	0.0794	0.0314	135	0.0073
IX	10^{-3}		0.2480	0.2497	0.0898	57	0.0704
	10^{-2}		1.2677	0.7566	0.1860	17	0.6071

(1) standard deviation; (2) estimate - real value; (3) relative bias =(abs.value(bias)/real)×100%

(4) % improvement to plan I= ((plan Y - plan I)/(plan I)) × 100%

Table 7: Simulation results (1000 samples) - percentiles t_p of a Weibull with $\alpha = 0.02$ and $\delta = 1.6$

plan	p	true value	estimate	SD(1)	bias(2)	RB(3) (%)	MSE
N=357	10^{-6}	0.0089	0.0170	0.0211	0.0081	91	0.0005
nw=51	10^{-5}	0.0375	0.0590	0.0590	0.0215	57	0.0039
np=7	10^{-4}	0.1581	0.02116	0.1640	0.0535	34	0.0298
I	10^{-3}	0.6670	0.7848	0.4417	0.1178	18	0.2090
	10^{-2}	2.8205	3.0236	1,0872	0.2030	7	1.2232
N=714	10^{-6}		0.0120	0.0105 (-50)(4)	0.0032 (-60)	35 (-62)	0.0001
nw=51	10^{-5}		0.0461	0.0324 (-45)	0.0086 (-60)	23 (-60)	0.0011
np=14	10^{-4}		0.1795	0.0984 (-40)	0.0213 (-60)	13 (-62)	0.0101
II	10^{-3}		0.7122	0.2854 (-35)	0.0452 (-62)	7 (-61)	0.0835
	10^{-2}		2.8896	0.7425 (-46)	0.0691 (-66)	2 (-71)	0.5560
N=280	10^{-6}		0.0221	0.0343	0.0132	149	0.0014
nw=40	10^{-5}		0.0706	0.0879	0.0331	88	0.0088
np=7	10^{-4}		0.2348	0.2258	0.0767	48	0.0057
III	10^{-3}		0.8205	0.5691	0.1536	23	0.3474
	10^{-2}		3.0446	1.3190	0.2241	8	1.7899
N=560	10^{-6}		0.0143	0.0165 (-22)	0.0054 (-33)	61 (-33)	0.0003
nw=40	10^{-5}		0.0518	0.0470 (-20)	0.0143 (-33)	38 (-33)	0.0024
np=14	10^{-4}		0.1925	0.1328 (-19)	0.0343 (-36)	22 (-35)	0.0188
IV	10^{-3}		0.7367	0.3622 (-18)	0.0698 (-41)	10 (-44)	0.1360
	10^{-2}		2.9184	0.8875 (-18)	0.0979 (-52)	3 (-57)	0.7972
N=504	10^{-6}		0.0162	0.0230 (9)	0.0073 (-10)	82 (-10)	0.0006
nw=36	10^{-5}		0.0562	0.0600 (2)	0.0187 (-13)	50 (-12)	0.0040
np=14	10^{-4}		0.2019	0.1589 (-3)	0.0438 (-18)	28 (-18)	0.0272
V	10^{-3}		0.7538	0.4139 (-6)	0.0869 (-26)	13 (-28)	0.1788
	10^{-2}		2.9397	0.9784 (-10)	0.1191 (-41)	4 (-43)	0.9714
N=224	10^{-6}		0.0329	0.0697	0.0240	270	0.0054
nw=32	10^{-5}		0.0936	0.1475	0.0561	150	0.0025
np=7	10^{-4}		0.2818	0.3245	0.1237	78	0.1206
VI	10^{-3}		0.9063	0.7268	0.2393	36	0.5855
	10^{-2}		3.1596	1.5415	0.3391	12	2.4911
N=448	10^{-6}		0.0188	0.0269	0.0099	112	0.0008
nw=32	10^{-5}		0.0624	0.0703	0.0249	66	0.0056
np=14	10^{-4}		0.2150	0.1843	0.0569	36	0.0372
VII	10^{-3}		0.7749	0.4721	0.1078	16	0.2345
	10^{-2}		2.9485	1.0886	0.1280	5	1.2015
N=168	10^{-6}		0.0629	0.1627	0.0540	607	0.0294
nw=24	10^{-5}		0.1463	0.2943	0.1088	290	0.0984
np=7	10^{-4}		0.3668	0.4493	0.2087	132	0.3453
VIII	10^{-3}		1.0125	1.0530	0.3455	52	1.2282
	10^{-2}		3.1924	1.9606	0.3718	13	3.9820
N=336	10^{-6}		0.0288	0.0546	0.0200	224	0.0034
nw=24	10^{-5}		0.0839	0.1205	0.0464	124	0.0167
np=14	10^{-4}		0.2583	0.2737	0.1002	63	0.0849
IX	10^{-3}		0.8499	0.6225	0.1830	27	0.4210
	10^{-2}		3.0402	1.2932	0.2196	8	1.7206

(1) standard deviation; (2) estimated - real; (3) relative bias =(abs.value(bias)/real value)×100%

(4) % improvement to plan I= ((plan Y - plan I)/(plan I)) × 100%

Table 8: Simulation Results (1000 samples)- fraction of defectives at t_0 for a Weibull with $\alpha = 0.02$ and $\delta = 1.2$

plan	t_0	true value	estimate	SD(1)	bias(2)	RB(3) (%)	MSE
N=357	1	0.0019	0.0105	0.0064	0.0013	15	4.2×10^{-5}
nw=51	2	0.0208	0.0225	0.0109	0.0017	8	1.2×10^{-4}
np=7	4	0.0471	0.0489	0.0176	0.0018	4	3.1×10^{-4}
I	8	0.1050	0.1059	0.0254	0.0009	0.9	6.5×10^{-4}
	12	0.1651	0.1651	0.0286	7.2×10^{-6}	0.004	$8,2 \times 10^{-4}$
	16	0.2249	0.2243	0.0291	-0.0006	0.3	8.5×10^{-4}
	20	0.2832	0.2824	0.0284	-0.0008	0.3	8.1×10^{-4}
	24	0.3393	0.3386	0.0277	-0.0007	0.2	7.8×10^{-4}
	32	0.4431	0.4431	0.0300	4.6×10^{-5}	0.01	9.0×10^{-4}
N=714	1		0.0096	0.0043	0.0005	5	1.8×10^{-5}
nw=51	2		0.0213	0.0075	0.0005	2	5.7×10^{-5}
np=14	4		0.0473	0.0124	0.0002	0.4	2.0×10^{-4}
	8		0.1043	0.0180	-0.0007	0.7	0.0003
II	12		0.1637	0.0201	-0.0014	0.9	0.0004
	16		0.2231	0.0203	-0.0019	0.8	0.0004
	20		0.2812	0.0196	-0.0020	0.7	0.0004
	24		0.3374	0.0190	-0.0019	0.6	0.0004
	32		0.4417	0.0207	-0.0014	0.3	0.0004
N=280	1		0.0109	0.0077	0.0018	19	6.2×10^{-5}
nw=40	2		0.0229	0.0126	0.0021	10	1.6×10^{-4}
np=7	4		0.0491	0.0196	0.0019	4	4.0×10^{-4}
	8		0.1055	0.0270	0.0005	0.5	7.0×10^{-4}
III	12		0.1644	0.0292	-0.0006	0.4	9.0×10^{-4}
	16		0.2237	0.0291	-0.0012	0.5	8.0×10^{-4}
	20		0.2821	0.0290	-0.0011	0.4	8.0×10^{-4}
	24		0.3387	0.0309	-0.0006	0.2	0.0010
	32		0.4442	0.0420	0.0011	0.3	0.0018
N=560	1		0.0099	0.0055	0.0008	9	3.1×10^{-5}
nw=40	2		0.0217	0.0093	0.0009	4	8.8×10^{-5}
np=14	4		0.0477	0.0149	0.0005	1	2.2×10^{-4}
	8		0.1044	0.0207	-0.0005	0.5	4.3×10^{-4}
IV	12		0.1639	0.0223	-0.0012	0.7	5.0×10^{-4}
	16		0.2236	0.0218	-0.0014	0.6	$5.0 \times 10-4$
	20		0.2822	0.0211	-0.0011	0.4	4.0×10^{-4}
	24		0.3388	0.0218	-0.0005	0.1	5.0×10^{-4}
	32		0.4443	0.0294	0.0012	0.3	5.0×10^{-4}
N=504	1		0.0103	0.0058	0.0012	13	3.4×10^{-5}
nw=36	2		0.0222	0.0096	0.0014	7	9.4×10^{-5}
np=14	4		0.0484	0.0149	0.0013	3	2.2×10^{-4}
V	8		0.1053	0.0202	0.0003	0.3	4.1×10^{-4}
	12		0.1645	0.0214	-0.0006	0.4	4.6×10^{-4}
	16		0.2237	0.0213	-0.0012	0.5	4.5×10^{-4}
	20		0.2818	0.0220	-0.0014	0.5	4.9×10^{-4}
	24		0.3379	0.0250	-0.0014	0.4	6.3×10^{-4}
	32		0.4422	0.0364	-0.0009	0.2	0.0013

(1)standard deviation; (2) estimate - real value; (3) relative bias =(abs.value(bias)/real value)×100%

5 Motivating Example Revisited

The analysis of the data consisted of two parts:

1. for each storage condition and atribute separately, the model proposed in Section 3 was fitted and its adequacy was checked. Percentiles and fraction of defectives where then estimated;

2. the results are then compared whenever possible with the ones obtained by the usual approach of fitting a regression line relating the scores and the evaluation weeks. In this case, the shelf life is obtained by inverse regression.

5.1 Results obtained with the proposed model with the Weibull as the underlying distribution

"Room temperature and humidity" storage condition.

Table 9 presents the parameter estimates obtained with the proposed model fitted to the data coming from the "room temperature and humidity" storage condition.

Table 9: Parameter estimates for the condition "room temperature and humidity".

atribute	estimates			
	$\hat{\beta}_0$	$\hat{\gamma}$	$\hat{\alpha} = e^{\hat{\beta}_0}$	$\hat{\delta} = e^{\hat{\gamma}}$
odor	-3.96	0.28 (p=0.04)*	0.0191	1.3
flavor	-4.01	0.19 (p=0.04)*	0.0181	1.2
appearance	-3.95	0.62 (p=0.00003)*	0.0193	1.9

(*) p value from $H_0: \ \gamma = 0$ vs. $H_1: \ \gamma \neq 0$(all significant at 5%).

Figure 1 presents the plot of the hazard function for each atribute, for products stored at this condition. Note that the atribute "appearance" is the one which deteriorates slower than the others until the 20^{th} week. The other two atributes have similar behaviors.

Model adequacy was checked by using the test statistic C (Lee, 1992). No evidence of lack of fit was found (p values were 0.88, 0.69 and 0.63 with models fitted to "odor", "flavor" and "appearance" respectively).

Tables 10 and 11 present estimates of percentiles and fraction defectives calculated at specific time points respectively. The confidence intervals are wide, implying a low precision associated with the estimates.

If the food company decides to stablish as the shelf life the percentile 1% for this storage condition then, for each one of the atributes we have:

- **odor**: estimated shelf life is 1.6 weeks (95% confidence interval:[0.6; 4.2]);

Table 10: Percentiles estimates for each atribute ("room temperature and humidity")

atribute		estimates		
	p	$\hat{t}_p$ (weeks)	LI*	LS*
odor	10^{-6}	0.0015	0.00006	0.0355
	10^{-5}	0.0085	0.0006	0.1167
	10^{-4}	0.0485	0.0061	0.3835
	10^{-3}	0.2780	0.0612	1.2624
	10^{-2}	1.5986	0.6123	4.1739
	0.05	5.5004	3.0975	9.7675
	0.50	39.5969	33,3582	47.0024
	0.63	52.0520	41,5958	65.1368
flavor	10^{-6}	0.0006	0.00002	0.0216
	10^{-5}	0.0039	0.0002	0.0774
	10^{-4}	0.0262	0.0025	0.2773
	10^{-3}	0.1778	0.0318	0.9950
	10^{-2}	1.2079	0.4066	3.5879
	0.05	4.6770	2.4495	8.9304
	0.50	40.6592	33.5142	49.3274
	0.63	54.8639	42.3966	70.9974
appearance	10^{-6}	0.0308	0.0036	0.2634
	10^{-5}	0.1061	0.0181	0.6231
	10^{-4}	0.3661	0.0909	1,4744
	10^{-3}	1.2635	0.4572	3.4920
	10^{-2}	4.3703	2.3004	8.3025
	0.05	10.5017	7.1677	15.3863
	0.50	42.6050	37.1621	48.8452
	0.63	54.8639	42.3966	70.9974

(*) 95% confidence limits.

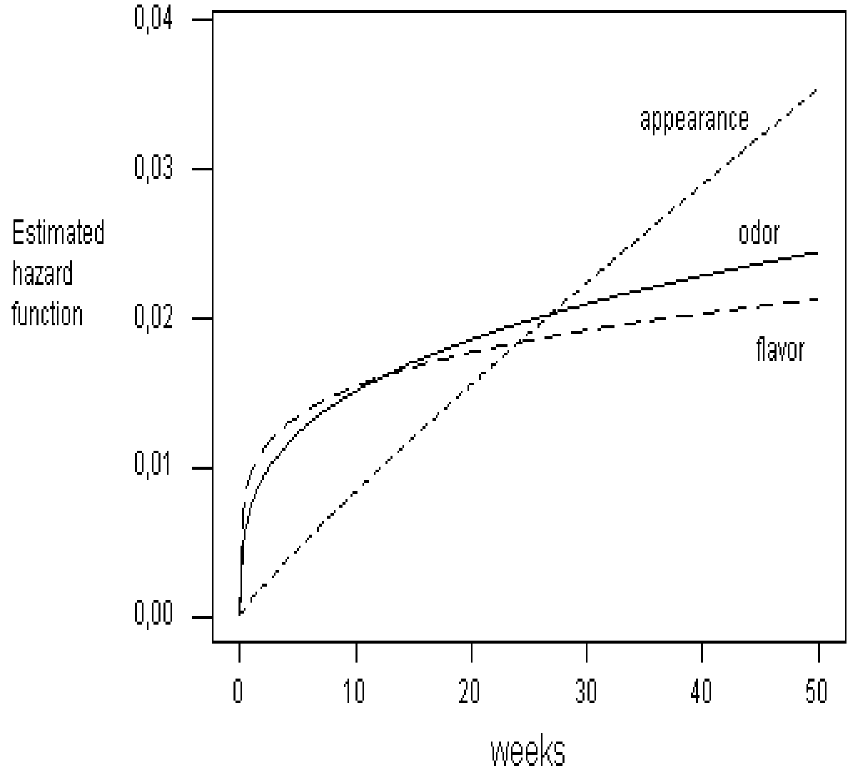

Figure 1: Estimated hazard function for each atribute ("room temperature and humidity").

- **flavor**: shelf life is 1.2 weeks ([0.4;3.6]);
- **appearance**: shelf life is 4.4 weeks ([2.3; 8.3]).

If we take into account the three atributes together, the shelf life should be 1.2 weeks (the smallest one).

The food company estipulates a "sell by"date of 48 weeks. since this date is supposed to allow a reasonable period of home storage and use, it is fair to say that the shelf life is greater than this "sell by" date.

One should keep in mind that the data that has been analysed is related only to the "sensory quality characteristics" of the product.

Taking into account the results presented in Table 11, at after 48 weeks of storage, we expect to have:

- **odor**: fraction defective estimate of 59% ([49;69]). In other words, it is expected that 59% of the products will experience a decrease in the quality level regarding this atribute;
- **flavor**: 57% ([48;67]);
- **appearance**: 58% ([48;69]).

Table 11: Fraction of defective estimates at t_0 weeks ("room temperature and humidity" storage condition)

atribute		estimates		
	t_0 (weeks)	fraction defectives	LI*	LS*
odor	1	0.0054	0.0013	0.0231
	2	0.0134	0.0042	0.0427
	4	0.0331	0.0138	0.0786
	8	0.0806	0.0448	0.1430
	12	0.1337	0.0875	0.2015
	16	0.1892	0.1379	0.2566
	20	0.2439	0.1924	0.3099
	24	0.3010	0.2475	0.3630
	32	0.4075	0.3477	0.4733
	36	0.4574	0.3903	0.5301
	40	0.5046	0.4283	0.5862
	44	0.5491	0.4626	0.6401
	48	0.5908	0.4937	0.6905
	52	0.6295	0.5224	0.7367
flavor	1	0.0080	0.0020	0.0315
	2	0.0183	0.0061	0.0545
	4	0.0416	0.0182	0.0937
	8	0.0933	0.0537	0.1595
	12	0.1474	0.0990	0.2165
	16	0.2019	0.1499	0.2688
	20	0.2555	0.2028	0.3188
	24	0.3075	0.2546	0.3684
	32	0.4052	0.3456	0.4708
	36	0.4505	0.3837	0.5232
	40	0.4932	0.4174	0.5747
	44	0.5334	0.4477	0.6242
	48	0.5711	0.4753	0.6708
	52	0.6063	0.5006	0.7138
appearance	1	0.0006	0.00009	0.005
	2	0.0023	0.0005	0.0119
	4	0.0085	0.0025	0.0291
	8	0.0305	0.0130	0.0704
	12	0.0636	0.0341	0.1170
	16	0.1061	0.0665	0.1672
	20	0.1563	0.1096	0.2201
	24	0.2122	0.1615	0.2759
	32	0.3344	0.2770	0.4000
	36	0.3976	0.3329	0.4698
	40	0.4601	0.3845	0.5429
	44	0.5209	0.4320	0.6162
	48	0.5790	0.4757	0.6763
	52	0.6336	0.5163	0.7504

(*)95% confidence limits.

Chamber 1 (30 ^oC; 80% relative humidity)and Chamber 2(temperature controlled at 37 ^oC).

In this situation, it was assumed that T_{ij}, the time to failure of the j^{th} unit evaluated at τ_i (fixed evaluation time) has a Weibull (α_j; δ) distribution ($\delta \geq 1$) where:

$$\begin{aligned} \alpha_j &= \exp(X_j^0\beta_0 + X_j^1\beta_1) \\ \text{and } \delta &= \exp(\gamma) \text{ for all } j = 1, 2, ..., n_i \text{ and } i = 1, 2, ..., k \end{aligned}$$

with: $k = 18 + 36$ (chambers 1 and 2) and

$$X_j^1 = \begin{cases} 0 & \text{if the } j^{th} \text{ unit was stored in chamber 2} \\ 1 & \text{if the } j^{th} \text{ unit was stored in chamber 1} \end{cases}$$

Here, $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma)'$. Tables 12 and 13 present the results of the model fitting.

Table 12: Parameter estimates for the model with a "dummy" variable for chamber 1 and 2

	estimates			
atribute	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$(\hat{\delta})$
odor	-3.50	0.68	0.48	(1.6)
		(p=0)$^{(*)}$	(p=0)$^{(*)}$	
sabor	-3.33	0.61	0.31	(1.4)
		(p=0.000006)$^{(*)}$	(p=0.02)$^{(*)}$	
aspecto	-3.76	0.95	0.68	(2.0)
		(p=0)$^{(*)}$	(p=0.00002)$^{(*)}$	

(*) p-value for $H_0 : \gamma = 0$ (significant at 5% level).

Table 13: Parameter estimates of the Weibull distribution (for each condition)

	atribute					
	odor		flavor		appearance	
condição	$\hat{\alpha}^*$	$\hat{\delta}$	$\hat{\alpha}^*$	$\hat{\delta}$	$\hat{\alpha}^*$	$\hat{\delta}$
chamber 1	0.0302	1.6	0.0358	1.4	0.0233	2
chamber 2	0.0596	1.6	0.0659	1.4	0.0602	2

(*) $\hat{\alpha} = exp(\hat{\beta}_0 + \hat{\beta}_1\, x)$ where x=0 ou 1.

Figure 2 presents the estimated hazard functions for the atribute "flavor", for each storage condition. Model adequacy was checked for each condition/atribute by using the test statistic C (Lee, 1992). Is all cases there is no evidence to conclude that the models are inadequate (smallest p value = 0.19).

Tables 14 and 15 present the percentiles and fraction defectives estimates for each condition.

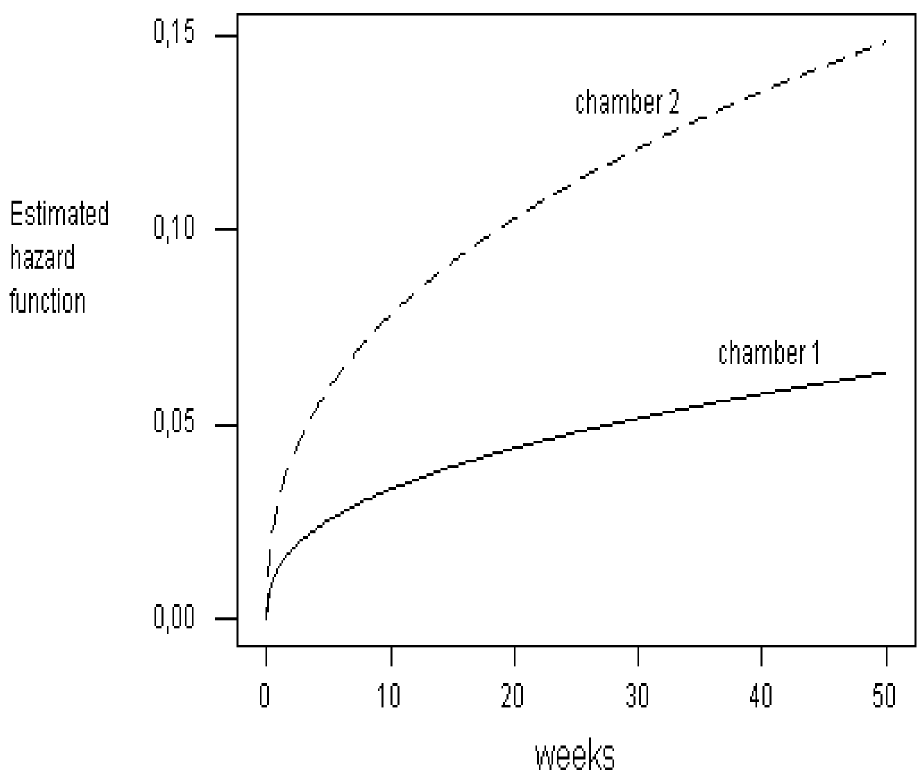

Figure 2: Hazard function plot ("flavor") for products stored in chambers 1 and 2

Table 14: Percentiles estimates ($\hat{t}_p$ weeks) (storage conditions: chambers 1 and 2)

		atributes		
	p	odor	flavor	appearance
chamber 1	10^{-6}	0.0064	0.0010	0.0381
		(0.0006; 0.0669)*	(0.00006; 0.0157)	(0.0048; 0.3045)
	10^{-5}	0.0265	0.0054	0.1229
		(0.0038; 0.1847)	(0.0005; 0.0542)	(0.0224; 0.6754)
	10^{-4}	0.1100	0.0301	0.3966
		(0.0237; 0.5103)	(0.0049; 0.1865)	(0.1050; 1.4994)
	10^{-3}	0.4568	0.1663	1.2810
		(0.1478; 1.4115)	(0.0430; 0.6428)	(0.4921; 3.3345)
	10^{-2}	1.9017	0.9214	4.1444
		(0.9220; 3,9228)	(0.3813; 2.2263)	(2.3022; 7.4610)
	0.05	5.2093	3.0889	9.4999
		(3,3393; 8.1265)	(1.7724; 5.3834)	(6.7711; 13.3286)
	0.50	26.3335	21.4991	35.7320
		(22.0239; 33.9882)	(18.0625; 25.2091)	(28.5776; 51.4514)
	0.63	32.5615	27.8813	42.9463
		(27.1043; 39,1175)	(23.3141; 33.3431)	(33.7206; 54.6962)
chamber 2	10^{-6}	0.0032	0.0005	0.0147
		(0.0003; 0.0349)	(0.00003; 0.0086)	(0.0017; 0.1305)
	10^{-5}	0.0135	0.0030	0.0474
		(0.0019; 0.0964)	(0.0003; 0.0297)	(0.0078; 0.2894)
	10^{-4}	0.0559	0.0163	0.1530
		(0.0117; 0.2666)	(0.0026; 0.1023)	(0.0364; 0.6426)
	10^{-3}	0.2320	0.0901	0.4941
		(0.0728; 0.7388)	(0.0230; 0.3534)	(0.1708; 1.4291)
	10^{-2}	0.9657	0.4991	1.5986
		(0.4524; 2.0613)	(0.2026; 1.2298)	(0.7997; 3.1955)
	0.05	2.6452	1.6733	3.6644
		(1.6284; 4.2972)	(0.9343; 2.9968)	(2.3611; 5.6872)
	0.50	13.3435	11.6793	13.8298
		(11.6882; 14.7259)	(9.3733; 14.2473)	(11.8921; 15.7077)
	0.63	16.5344	15.1035	16.5656
		(13.2943; 20.5642)	(11.9220; 19.1341)	(13.7009; 20.0293)

(*) 95% confidence interval.

Table 15: Fraction of defectives (at t_0 weeks) estimates (storage conditions: chambers 1 and 2)

		atributo		
	t_0	odor	sabor	aspecto
chamber 1	1	0.0035	0.0112	0.0006
		(0.0008; 0.0153)*	(0.0035; 0.0351)	(0,00008; 0.0047)
	2	0.0184	0.0282	0.0024
		(0,0035; 0.0338)	(0,0115; 0.0682)	(0,0005; 0.0119)
	4	0.0329	0.0701	0.0093
		(0.0144; 0.0743)	(0.0370; 0.1307)	(0.0029; 0.0300)
	8	0.0976	0.1688	0.0359
		(0.0583; 0.1609)	(0.1143; 0.2454)	(0,0169; 0.0755)
	12	0.1795	0.2733	0.0780
		(0.1269; 0.2505)	(0.2105; 0.3504)	(0.0463; 0.1299)
	16	0.2702	0.3753	0.1331
		(0.2106; 0.3428)	(0.3096; 0.4498)	(0.0910; 0.1925)
	20	0.3636	0.4703	0.1987
		(0.2977; 0.4391)	(0.4004; 0.5460)	(0.1472; 0.2652)
	24	0.4550	0.5562	0.2716
		(0.3789; 0.5387)	(0.4783; 0.6374)	(0.2078; 0.3519)
	32	0.6197	0.6979	0.4275
		(0.5141; 0.7260)	(0.5998; 0.7908)	(0.3237; 0.5485)
chamber 2	1	0.0106	0.0253	0.0040
		(0.0032; 0.0351)*	(0.0097; 0.0651)	(0.0008; 0.0206)
	2	0.0321	0.0631	0.0155
		(0.0131; 0.0775)	(0.0311; 0.1259)	(0.0046; 0.0517)
	4	0.0953	0.1529	0.0591
		(0.0527; 0.1691)	(0.0957; 0.2395)	(0.0267; 0.1284)
	8	0.2645	0.3447	0.2117
		(0.1892; 0.3624)	(0.2610; 0.4454)	(0.1403; 0.2123)
	12	0.4468	0.5177	0.4110
		(0.3480; 0.5593)	(0.4163; 0.6276)	(0.2128; 0.5240)
	16	0.6105	0.6586	0.6049
		(0.4864; 0.7366)	(0.5396; 0.7743)	(0.4754; 0.7373)
	20	0.7414	0.7658	0.7630
		(0.5977; 0.8659)	(0.6349; 0.8764)	(0.6055; 0.8923)
	24	0.8374	0.8436	0.8725
		(0.6862; 0.9420)	(0.7092; 0.9384)	(0.7071; 0.9684)
	32	0.9446	0.9351	0.9734
		(0.8115; 0.9934)	(0.8141; 0.9882)	(0.8449; 0.9991)

(*) 95% confidence interval.

Here, the results seem to indicate that units stored in chamber 2 (37^oC) had the atributes deteriorated faster than those products stored in chamber 1.

So, as we have done in the case of the storage condition "room temperature and humidity", if the food company decides to use as a shelf life the percentile 1% then (see Table 14):

- for units stored in chamber 1 (30^oC and 80%):
  - "odor": 1.9 weeks (13 days) ([0.9 weeks; 3.9 weeks]);
  - "flavor": 0.9 weeks (6 days) ([0.4 weeks; 2.2 weeks]) and
  - "appearance": 4.1 weeks (29 days) ([2.3 weeks; 7.5 weeks]);
- for units stored in chamber 2(37^oC):
  - "odor": 1 week (7 days) ([0.5 weeks; 2.1 weeks]);
  - "flavor": 0.5 weeks (3.5 days) ([0.2 weeks; 1.2 weeks]) and
  - "appearance": 1.6 weeks (11 days) ([0.8 weeks; 3.2 weeks]);

5.2 The Usual Approach

"Room Temperature and Humidity" condition.

We have already mentioned that the usual approach to estimate the shelf life is to fit a linear regression model. For the real data set under study we have:

$$y_{ij} = \beta_0 + \beta_1 week_i + \varepsilon_{ij} \quad i = 1, ..., 51 \ (week) \ \ j = 1, 2, .., n_i \qquad (16)$$

where:

- y_{ij} is the score of the j^{th} unit evaluated at the i^{th} week;
- $week_i$ is the independent variable representing the evaluation weeks. Therefore, $week_1$ =1st.week, ..., $week_{51} = 51^{th}$ week;
- n_i is the number of replicates (ou judges) at the i^{th} week.

The results of the fitted model are in Table 16.

There is no indication of lack of fit (5th column of Table 16). So, we conclude that, in fact, the scores decrease linearly in magnitude with the weeks.

The third column of Table 16 presents the estimates (using inverse regression) of the shelf life. In fact, those are the weeks in which the mean score reaches the value three (3).

Therefore, the estimated shelf life is 49, 51 e 52 weeks respectively for the atributes "odor", "flavor" and "appearance" respectively. 95% confidence intervals were also constructed (Draper and Smith, 1981 - page 49).

Table 16: Simple linear regression model fitted to the data comming from the storage condition "room and temperature humidity"

atribute	estimates			other measures	
	$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1$	week when score=3$^{(2)}$	R^2	lack of fit test$^{(3)}$
odor	5.66 (p=0)	-0.0544 (p=0)	49 (44;56)	24%	0.65 (p=0.95)
sabor	5.46 (p=0)	-0.0485 (p=0)	51 (46;59)	22%	0.66 (p=0.95)
aspecto	6.00 (p=0)	-0.0582 (p=0)	52 (48;56)	38%	1.25 (p=0.15)

(1) parameter estimate and p-value for H_0:$\beta_u = 0$; (2) obtained by using inverse regression; (3) lack of fit test: H_0: no lack of fit

Results for Chamber 1 (30^oC and 80%)and Chamber 2 (37^oC).

A linear regression model was also fitted to the data coming from units stored in chambers 1 and 2. The results are presented in Table 17.

Table 17: Simple linear regression model fitted to chambers 1 and 2.

storage	atribute	estimates			other measures	
		$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1$	semana quando nota=3$^{(2)}$	R^2	adequação do modelo$^{(3)}$
chamber 1	odor	5.54 (p=0)	-0.0761 (p=0)	33 (30;38)	24%	0.65 (p=0.93)
	flavor	5.47 (p=0)	-0.0944 (p=0)	26 (23;29)	31%	0.69 (p=0.89)
	appearance	5.98 (p=0)	-0.0632 (p=0)	47 (41;57)	23%	1.02 (p=0.44)
chamber 2	odor	6.21 (p=0)	-0.1970 (p=0)	16 (15;18)	43%	1.34 (p=0.19)
	flavor	5.60 (p=0)	-0.171 (p=0)	15 (14;17)	36%	0.70 (p=0.78)
	appearance	5.77 (p=0)	-0.166 (p=0)	17 (15;19)	34%	2.38 (p=0.005)$^{(4)}$

(1) parameter estimate and p-value for H_0:$\beta_u = 0$; (2) obtained by using inverse regression; (3) lack of fit test:H_0: no lack of fit; (4) significant at 5%

Note that for Chamber 1, there is no indication of lack of fit.

The point estimates for the shelf life are 33, 26 e 47 weeks for "odor", "flavor" and "appearance" respectively.

On the other hand, for Chamber 2 there is an indication of lack of fit for the atribute "appearance" (p=0.005). The usual residual plots were constructed to identify possible violations on the model assumptions. They all indicated problems with the homocedasticity assumption. The shelf life point estimates are 16 and 15 weeks for the atributes "odor" and "flavor" respectively. These values are smaller than the ones obtained for Chamber 1.

5.3 Results Comparison

Table 18 presents the shelf life estimates for "room temperature and humidity" condition, according to each of the two approaches: simple linear regression and the proposed model.

Table 18: Shelf life estimates obtained according to each approach ("room temperature and humidity" condition).

	approach		
	linear regression	proposed model	
atribute	shelf life (x_0) x_0 = week when mean score(or median) =3	shelf life (week) [1]	fraction defectives at x_0[2] (%)
odor	49 (44; 56)	$t_{0.50}$=40 (33; 47)	60 (50; 70)
flavor	51 (46; 59)	$t_{0.50}$=41 (34; 49)	60 (49; 70)
appearance	52 (48; 56)	$t_{0.50}$=43 (37; 49)	63 (52; 75)

(1) values in Table 10; (2) calculated with α and δ estimates in Table 9.

The values shown in column 1 of Table 18 were calculated by inverse regression. In other words, they were obtained by making "Score=3" in the fitted regression line (Score=a+b * Week) and then solve to find the week value.

Let us assume for a moment that the usual model assumptions (normality and homocedasticity) are valid. Then, since the mean and median are the same in the Normal distribution it is fair to say that, the estimated fraction of defectives at weeks 49 (odor), 51 (flavor) and 52 (appearance) are all 50%.

The third column of Table 18 shows the fraction of defective estimates at weeks 49, 51 and 52, calculated with the proposed model. In other words, using the proposed approach, the fraction of defectives at weeks 49, 51 and 52 are 60%, 60% and 63% respectively.

Now, in order to have comparable results, the second column presents the shelf life estimates, calculated with the proposed approach. There, it was assumed that the company has chosen the percentiles 50% as the shelf life values.

As Tables 19 and 20 present similar results calculated for chambers 1 and 2 respectively.

Table 19: Shelf life estimates obtained according to each approach (Chamber 1: [30^oC; 80%]).

	approach		
	linear regression	proposed model	
atribute	shelf life (x_0) x_0 = week when mean score (or median) =3	shelf life (week) [1]	fraction defectives at ${x_0}^{(2)}$ (%)
odor	33 (30; 38)	$t_{0.50}$=26 (22; 34)	63 (53; 75)
flavor	26 (23; 29)	$t_{0.50}$=21 (18; 25)	61 (51; 68)
appearance	47 (41; 57)	$t_{0.50}$=36 (29; 51)	65 (51; 86)

(1) values in Table 14; (2) calculated with α and δ estimates in Table 13.

Table 20: Shelf life estimates obtained according to each approach (Chamber 2: [37^oC]).

	approach		
	linear regression	proposed model	
atribute	shelf life(x_0) x_0 = week when mean score (or median) =3	shelf life (week) [1]	fraction defectives at ${x_0}^{(2)}$ (%)
odor	16 (15; 18)	$t_{0.50}$=13 (12; 15)	62 (49; 74)
flavor	15 (14; 17)	$t_{0.50}$=12 (9; 14)	63 (51; 74)
appearance	17 (15; 19)	$t_{0.50}$=14 (12; 16)	64 (51; 78)

(1) values in Table 14; (2) calculated with α and δ estimates in Table 13.

6 Discussion

The two approaches discussed in this article have both strong and weak points.

The shelf life estimation via linear regression models relating the scores and the (fixed) evaluation times has the advantage of been very easy to implement and the results are easy to interpret. In addition, there is a great number of comercial softwares available which deal with regression analysis and model checking.

One of the drawbacks associated with this approach is the model assumptions. The data analysis and the analysis of similar kind of data indicate that, at least, the homocedasticity assumption is not valid for data arising from these situations. In other words, the panelists all seam to “agree” at the begining of the study. As the time goes on and the products’ atributes start to show some kind of deterioration, the variability on the scores increases. But we believe that the main disadvantage of this approach is the difficulty to estimate percentiles and fraction defectives at choosen time points. This information is crucial to support managers decisions, in particular, the ones associated with determining the shelf life. Moreover, since the shelf life estimate is obtained in this case by inverse regression, it is very difficult to incorporate covariates in the model and then construct confidence intervals for such quantity.

On the other hand, the approach proposed in this article incorporates both the information contained in the "failed" and "unfailed" units. Moreover, it provides estimates of percentiles and fraction of defectives. Covariates are easily incorporated in the model if necessary.

Some limitations we can recognize in our approach. First, as in most of the parametric statistical methods, if one chooses an inapropriate underlying distribution, the results (as with the regression approach) could be totally unreliable. Another limitation has to do with the dichotomization implemented. We are not using all the "score information" available.

In both approaches, the shelf life estimate will be highly affected by the the cut-off point chosen by the food company.

We also see some issues for further researches:

- one possibility is to try to model two atributes together, in other words, to work with the joint distribution of their time to failure;
- in our approach we did not question the "scores" given by the panelists (we assumed that they "never fail"). One possibility is to incorporate "classification errors".

To summarize, we believe that our approach will provide a much needed practical approach to setting realistic and economic "open dates" useful for producing manufactured products subject to sensory evaluations.

Acknowledgments

The authors would like to thank CAPES/Brazil for the financial support of this research.

References

Cox, D.R.; Hinkley, D.U.(1974).*Theoretical Statistics.* London: Chapman and Hall.

Draper, N.; Smith, H.(1998) *Applied Regression Analysis.* 3rd.ed. New York: Wiley.

Gacula, M.C. Jr.(1975). "The design of experiments for shelf life study". *Journal of Food Science*,40, pp. 399-403.

Gacula, M.C. Jr.; Kubala, J.J.(1975).“Statistical models for shelf life failures”. *Journal of Food Science*, 40, pp.404-409.

Gacula, M.C. Jr.; Singh,J. (1984).*Statistical Methods in Food and Consumer Research.* New York: Academic Press.

Jay, James M.(1992). *Modern Food Micro-Biology.* 4th.ed. New York: Chapman & Hall.

Kalbfleisch, J.D.; Prentice, R.L.(1992). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Lawless,J.F.(1980).*Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Meeker, W.Q.; Escobar, L.A. (1998). *Statistical Methods for Reliability Data.* New York: Wiley.

MINITAB, Statistical Software, v.12, Minitab Inc., 1998.

SAS, Statistical Analysis System. Windows version, 1994.

Seber, G.A.F.; Wild, C.J. (1989). *Nonlinear Regression.* New York: Wiley.