Daily Counting of Manufactured Units sent for Quality Control: a Bayesian Approach

Jorge Alberto Achcar*+, Claudio Luis Piratelli*, Renata Regina Sandrim*, José Luiz Garcia Hermosilla*

*Master Program in Production Engineering, University Center of Araraquara, Araraquara, Brazil.
+ Department of Social Medicine, University of São Paulo, Brazil

Email: achcar@frmp.usp.br, clpiratelli@uniara.com.br, renatasandrim@hotmail.com, jlghermosilla@hotmail.com

Abstract

In this paper, it is presented a statistical modeling approach for daily counting data of units arriving at a quality control department of a food industry. Different regression Poisson models were considered to analyze the selected data, under a Bayesian approach. The counting was made as a diagnostic for the utilization rate of the quality control operators in the quality control department, using simulation methods. The discovering of important factors that affect these variability was essential for the industrial manager to take decisions in terms of reallocation of workers in the department (due to small mean rate of utilization) and for the programming of the statistical analyses (quality control), which implies in better planning of the productive capacity, without any significative effect on the lead-time of the quality analyses.

Keywords: Poisson regression; Bayesian analysis; Markov Chain Monte Carlo methods.

1 Introduction

Great variability in the daily counting of manufactured units in a production line it is common in industries. The statistical modeling of these data are useful as a diagnostic to the production systems (as an example, in measuring the performance indicators using queue modeling theory), inference or simulation, especially for decisions in investment, production programming, allocation of the productive capacity, among many others). The discovering of these factors affecting the variability can be of great interest for industrial engineers and industrial managers.

In practical work, in general there are two kinds of data which are fundamental for the modeling of productive systems: the times between the arrivals of manufactured units or the daily counting of the manufactured units.

In this paper, it is considered the daily counting data of manufactured units sent to quality control of a food industry located in the São Paulo State, Brazil in a period of 30 non-consecutive days. This data set is presented in Table 1. Figure 1 shows the plot of the observed number of units (daily counting) against days, where it is possible to observe the great variability of the daily counting.

The main goal of this paper is to analyze the daily counting data (arrivals to quality control) of Table 1, using Poisson regression models that assume some covariates like: days where it was collected the data, and seasonality factors (specific days of the week).

The step of identifying probability distributions of times and its parameters, is often dealt superficially in papers about Discrete Event Simulation Modeling - see, for example, Sakurada and Miyake (2009). In Queueing Theory, by other hand, the identification of distributions and possible covariates that affect data has been focus of several studies, for example, Shin and Pearce (1998), Ausín et al. (2004), Ausín et al. (2008), Wu and Hui (2008) and others.

Bayesian methods have been extensively used in many applied areas as Business Administration, Economy and Industrial Engineering: Motta (1997), Quinino and Bueno Neto (1997), Ramirez Pongo and

Bueno Neto (1997), Droguett and Mosleh (2006), Cavalcante and Almeida (2006), Kalatzis et al. (2006), Moura et al (2007), Ferreira et al. (2009), Barossi-Filho et al (2010) and Freitas et al. (2010).

In this work, the inferences of interest were obtained using Bayesian methods. The posterior summaries of interest were obtained via MCMC (Markov Chain Monte Carlo) simulation methods, as the popular Gibbs sampling algorithm (Gelfand and Smith, 1990) or the Metropolis–Hastings algorithm when the conditional posterior distributions required for the Gibbs sampling algorithm do not have standard parametrical forms (see for example, Chib and Greenberg, 1995).

Table 1: Daily counting data of manufactured units send to quality control

Observation i	Date	Number of Units	Observation i	Date	Number of Units
1	2011.04.04*	68	16	2011.04.27	123
2	2011.04.05	99	17	2011.04.28	99
3	2011.04.06	125	18	2011.04.29	108
4	2011.04.07	117	19	2011.05.02*	91
5	2011.04.08	84	20	2011.05.03	91
6	2011.04.11*	73	21	2011.05.30*	58
7	2011.04.12	100	22	2011.05.31	59
8	2011.04.13	103	23	2011.06.01	70
9	2011.04.14	91	24	2011.06.02	75
10	2011.04.15	88	25	2011.06.03	58
11	2011.04.18*	73	26	2011.06.06*	68
12	2011.04.19	93	27	2011.06.07	90
13	2011.04.20	95	28	2011.06.08	107
14	2011.04.25*	95	29	2011.06.09	102
15	2011.04.26	113	30	2011.06.10	90

(observation with (*) is related to Mondays). Source: own.

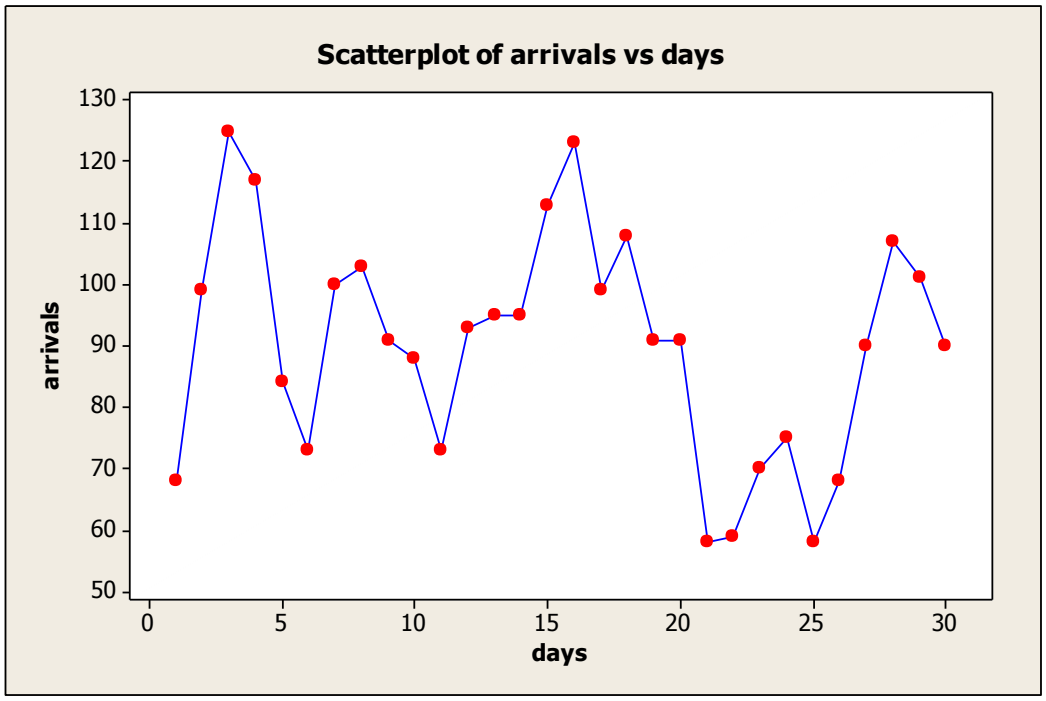

Figure 1: Daily counting (arrivals) versus days.

The paper is organized as follows: in section 2 we present some Poisson regression models; in section 3, we present a statistical analysis of the data presented in Table1; finally, in section 4, we present a discussion on the obtained results and in section 5, we present some concluding remarks.

2 Statistical modelling

Let Y_i be a random variable with a Poisson distribution given by,

$$P\,(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \tag{1}$$

where y_i = 0, 1, 2, ... denotes the daily number of units (arrivals) for quality control in the i-th day, i = 1, 2, ..., 30. Observe that the mean and the variance of the Poisson distribution (1) are equal to λ_i.

Different Poisson regression models can be considered to analyze the data set presented in Table 1, given as follows:

Model 1: Consider λ_i. given in (1), by,

$$\lambda_i = exp\,(\beta + \beta_1 X_i)\,, \tag{2}$$

where X_i denotes the i-th day (X_i = 1, 2, ...,30). In this case, it is assumed that days could influence the daily counting of units sent to quality control. The formulation (2) guarantees λ_i to be positive, for i = 1, 2, ..., 30.

Model 2: In this case, there is the addition of the quadratic term X_i^2 in model (2), that is,

$$\lambda_i = exp\left(\beta_0 + \beta_1 X_i + \beta_2 X_i^2\right)\,, \tag{3}$$

where X_i is defined in (2), i = 1,2, ..., 30.

Model 3: In this model, also there is the presence of an autoregressive effect of the daily counting y_{i-1} in the modeling for λ_i , that is, a first order autoregressive model given by ,

$$\lambda_1 = exp(\beta_0 + \beta_1 X_1),$$
$$\lambda_i = exp\,(\beta_0 + \beta_1 X_i + \beta_2 y_{i-1})\,, \tag{4}$$

where X_i is given in (2), and y_{i-1} is the number of units (arrivals for the quality control) in the (i-1)-th day, i = 2, 3, ..., 30.

Model 4: Consider the introduction of another autoregressive term in λ_i, given by,

$$\lambda_1 = exp\,(\beta_0 + \beta_1 X_1),$$
$$\lambda_2 = exp(\beta_0 + \beta_1 X_2 + \beta_2 y_1),$$
$$\lambda_i = exp(\beta_0 + \beta_1 X_i + \beta_2 y_{i-1} + \beta_3 y_{i-2})\,, \tag{5}$$

where X_i and y_i are defined, respectively in (2) and (4), for i = 3,4, ..., 30.

Model 5 : Consider λ_i, in (1), given by,

$$\lambda_1 = exp\,(\beta_0 + \beta_1 X_1),$$
$$\lambda_2 = exp\,(\beta_0 + \beta_1 X_2 + \beta_2 y_1),$$
$$\lambda_3 = exp\,(\beta_0 + \beta_1 X_3 + \beta_2 y_2 + \beta_3 y_1),$$
$$\lambda_i = exp(\beta_0 + \beta_1 X_i + \beta_2 y_{i-1} + \beta_3 y_{i-2} + \beta_4 y_{i-4}) \tag{6}$$

where i = 4, 5, ..., 30.

In the models defined by (1), (2), (3), (4), (5) and (6), we also could consider a seasonal effect. For the data presented in Table 1, it could be considered the effect of a specified week day, in our case, Mondays, where it seems to have a smaller number of units sent to quality control as compared to other week days.

In this way, let us define the indicator variable $z_i = 1$ when the i-th day is a Monday and $z_i = 0$ when the i-th day is not a Monday, for i = 1, 2, ..., 30.

Assuming the models defined above, the likelihood function for the vector of parameters θ associated to each model is given by,

$$L(\theta) = \prod_{i=1}^{30} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \tag{7}$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1)$ for model 1; $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2)$ for models 2 and 3;

$\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ for model 4 and $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ for model 5.

For a Bayesian analysis, let us assume the following prior distributions for the regression parameters β_j of the models:

$$\beta_j \sim N\left(a_j\,;\, b_j^2\right) \tag{8}$$

where $N\left(a_j\,;\, b_j^2\right)$ denotes a normal distribution with mean a_j and variance b_j^2, $j = 0, 1, 2, 3, 4$. Also it is assumed prior independence among the parameters of the models.

Combining the joint prior distribution (a product of normal distributions) with the likelihood function L(θ) given in (7), we obtain from the Bayes formula, the joint posterior distribution for θ (see for example, Box and Tiao, 1973).

The posterior summaries of interest are obtained using MCMC methods. A great simplification in the simulation of samples for the joint posterior distribution for θ is given using the WinBugs *software* (Spiegelhalter et al, 2003) which only requires the specification of the distribution for the data and the prior distributions for the parameters.

3 Analysis of the daily counting data of units sent to quality control

To analyze the quality control counting data (data set presented in Table 1), initially we consider the model 1, defined respectively by (1) and (2), with the following priors for β_0 and β_1: $\beta_0 \sim N(a_0, 10)$ and $\beta_1 \sim N(0,10)$, for fixed values of a_0. In this analysis, considering different values of a_0 we obtained similar results. With a fixed value for the variance of the normal distribution equals to 10, that is, a large value, we have approximately non-informative priors.

Using the WinBugs software, we first simulated a "burn-in-sample" of size 10,000 discarded to eliminate the effect of the initial values used in the Gibbs sampling algorithm; after this "burn-in-sample period", we simulated another 20,000 Gibbs samples for β_0 and β_1. From this sample, we selected a final sample of size 1000 , taking a sample chosen from every 20 simulated samples to have an approximately uncorrelated sample to be used to find the posterior summaries of interest. The obtained posterior summaries of interest (posterior mean, posterior standard deviation and 0.95% credible intervals) are given in Table 2. Convergence of the algorithm was monitored using standard traceplots of the simulated samples (see, for example, Paulino et al, 2003, or Gamerman, 1997).

In Table 2, we also have the posterior summaries of interest considering the models 2, 3, 4 and 5 assuming the same normal prior distributions N(0,10) for the regression parameters β_j , j = 1, 2, 3, 4.

In the simulation of Gibbs samples for the parameters of these models we have used the same steps in the WinBugs software used for model 1.

For selection of the best model to be fitted by the data, it was considered three discrimination criteria:

(a) Graphical verification of the fit: in this case, it was considered plots of the observed values y_i (counting of units sent to quality control in the i-th day) against days and plots of the fitted means (posterior mean of λ_i in (1)) considering the five proposed models.

(b) Use of the Bayesian discrimination method DIC (*Deviance Information Criterion*) introduced by Spiegelhalter et al (2002). This criterion for model selection is useful in the selection of models when samples of the joint posterior distribution for the parameters of the models were obtained by MCMC simulation methods. Smaller values of DIC indicates better models.

Table 1: Posterior summaries

Model	Parameters	Mean	S.D.	95% Credible Interval
Model 1	β_0	4.605	0.0379	(4.531 ; 4.677)
DIC = 299.84	β_1	-0.00653	0.00218	(-0.0106 ; -0.00216)
Model 2	β_0	4.617	0.06037	(4.500 ; 4.733)
	β_1	-0.00912	0.00892	(-0.0269 ; 0.00873)
DIC = 301.99	β_2	0.000082	0.00028	(-0.00047 ; 0.00066)
Model 3	β_0	4.211	0.08464	(4.034 ; 4.380)
	β_1	-0.00655	0.00225	(-0.01105 ; -0.00207)
DIC = 264.93	β_2	0.00445	0.000806	(0.00282 ; 0.006013)
	β_0	4.153	0.08615	(3.985 ; 4.3100)
Model 4	β_1	-0.00408	0.00225	(-0.0085 ; 0.0005)
DIC = 255.53	β_2	0.00772	0.001186	(0.0054 ; 0.0102)
	β_3	-0.00324	0.001004	(-0.0052 ; -0.0012)
	β_0	4.176	0.08917	(4.007 ; 4.349)
	β_1	-0.00366	0.00236	(-0.0084 ; 0.00088)
Model 5	β_2	0.00723	0.001286	(0.00486 ; 0.00993)
DIC = 256.30	β_3	-0.00204	0.00151	(-0.00521 ; 0.00091)
	β_4	-0.00109	0.001014	(-0.00312 ; 0.00089)

Source: own

(c) Determination of the sums of the absolute values of the differences between the observed values y_i with the fitted values (posterior means of Y_i), i = 1, 2, ..., 30, that is,

$$S_j = \sum_{i=1}^{30} | y_i - \tilde{\lambda}_i | \tag{9}$$

where y_i is the daily counting of units sent to quality control and $\tilde{\lambda}_i$ is the Monte Carlo estimate of the posterior mean for λ_i based on the 1000 simulated Gibbs samples.

In Figure 2, we have the graphs of the observed values y_i against days and the graphs of the estimated values ($\tilde{\lambda}_i$) against days, for i = 1, 2, ..., 30. We observe that models 4 and 5 have good fit for the counting data presented in Table 1. We also observe smaller values for DIC (see Table 2) assuming models 4 and 5, an indication of better models.

Considering the sums of the absolute values of the differences between the observed values y_i with the fitted values (9), we have: S_1 = 435.26 (model 1) ; S_2 = 437.42 (model 2); S_3 = 345.26 (model 3); S_4 = 314.19 (model 4) and S_5 = 299.19 (model 5). In this way, we conclude that models 4 and 5 are better fitted by the counting data of Table 1.

Assuming model 4 (a model with less parameters) , it was considered the presence of an indicator or "dummy" variable for week day , that is, Z_i = 1 for Mondays and Z_i = 0 for the other week days and the regression Poisson model ("model 6") given by,

$$\lambda_1 = \exp(\beta_0 + \beta_1 X_1 + \beta_4 Z_1)$$
$$\lambda_2 = \exp(\beta_0 + \beta_1 X_2 + \beta_2 y_1 + \beta_4 Z_2)$$
$$\lambda_i = \exp(\beta_0 + \beta_1 X_i + \beta_2 y_{i-1} + \beta_3 y_{i-2} + \beta_4 Z_i) \tag{10}$$

where i = 3, 4, ..., 30.

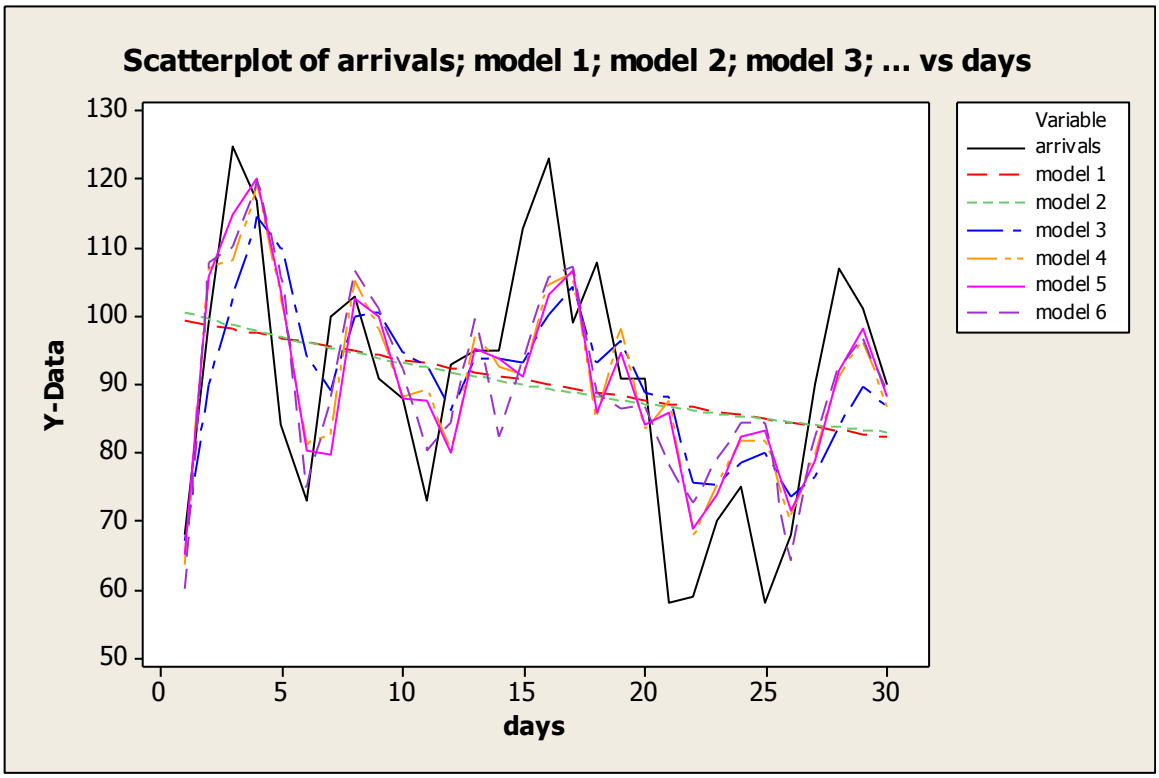

Figure 2: Daily counting (arrivals) versus days.

In Table 3, the posterior summaries of interest are given for the model 6 defined by (1) and (10) assuming normal prior distributions with variance equals to 10 and the same simulation steps considered for the other models using the WinBugs software.

Table 2: Posterior summaries ("model 6").

Parameters	Mean	S.D.	95% credible interval
β_0	4.245	0.09352	(4.060 ; 4.427)
β_1	- 0.00475	0.00240	(- 0.00912 ; 0.000197)
β_2	0.006532	0.00124	(0.00414 ; 0.00899)
β_3	- 0.00259	0.000984	(- 0.0044 ; -0.00068)
β_4	- 0.1462	0.05147	(- 0.2489 ; - 0.0452)

Source: Own

Assuming model 6, it was obtained DIC=249.12 and S6 = 304.33. We also observe in Figure 2, a good fit of the model for the counting data of Table 1.

4 Discussion of the obtained results

From the obtained results of Table 3, for "model 6", we observe that the covariates y_{i-1} (counting with a lag of one day) , y_{i-2} (counting with a lag of two days) and Z_i (weekday) present significative effects in the daily counting of units sent to quality control, since the zero value is not included in the 95% credible intervals for the regression parameters β_2 ,β_3 and β_4.

Since the Bayesian estimate of the parameter β_4 is negative ($\hat{\beta}_4$= - 0.1462), we observe that Mondays present smaller countings of units sent to quality control as compared to the other weekdays. This result could be of great interest for the industry.

Also it was possible to use the "model 6" to get predictions, from the fitted model (11):

$$\hat{\lambda}_i = \exp\,(4.245 - 0.00475\, X_i + 0.00653\, y_{i-1} - 0.00259\, y_{i-2} - 0.1462\, Z_i) \qquad (11)$$

In this way, considering the prediction for the future day (i=31), we have X_{31} = 31, y_{30} = 90, y_{29} = 101, Z_i = 1 (if the next day is a Monday), that is, $\widehat{\lambda_{31}}$ = 72.0716 (72 units); if Z_i = 0 (the next day is not a Monday) , we have $\widehat{\lambda_{31}}$ = 83.4177 (83 units).

Finally, in Table 4, we have the observed values (arrivals) and the fitted values for the means assuming each proposed model.

Table 3: Observed values and estimated means

Observation	Arrivals	Model-1	Model-2	Model-3	Model-4	Model-5	Model-6
1	68	99.42	100.50	67.21	63.62	65.12	60.20
2	99	98.77	99.54	90.14	107	105.90	108.00
3	125	98.12	98.65	102.80	108.30	114.90	110.10
4	117	97.48	97.79	114.70	119.20	120.30	119.90
5	84	96.84	96.97	109.90	102.60	103.70	105.90
6	73	96.20	96.16	94.25	81.33	80.41	74.97
7	100	95.57	95.39	89.17	82.77	79.83	87.48
8	103	94.95	94.64	99.89	105.20	102.50	106.80
9	91	94.33	93.91	100.60	98.22	99.92	101.10
10	88	93.71	93.20	94.71	88.31	88.08	92.32
11	73	93.10	92.52	92.84	89.34	87.71	80.33
12	93	92.49	91.85	86.28	80.04	79.94	84.53
13	95	91.89	91.21	93.69	97.64	95.21	99.64
14	95	91.29	90.58	93.91	92.54	93.91	82.48
15	113	90.69	89.98	93.30	91.57	91.16	94.45
16	123	90.10	89.39	100.40	104.80	103.20	105.70
17	99	89.52	88.82	104.30	106.40	106.60	107.20
18	108	88.93	88.27	93.13	85.22	85.80	88.93
19	91	88.36	87.73	96.31	98.32	94.78	86.36
20	91	87.78	87.21	88.70	83.41	84.15	86.90
21	58	87.21	86.72	88.13	87.77	85.94	78.13
22	59	86.65	86.23	75.61	67.81	68.78	72.57
23	70	86.09	85.77	75.46	75.73	73.80	79.15
24	75	85.53	85.33	78.71	81.83	82.36	84.41
25	58	84.98	84.90	79.96	81.73	83.11	84.35
26	68	84.43	84.49	73.67	70.26	71.65	64.12
27	90	83.88	84.11	76.51	79.87	78.98	82.36
28	107	83.34	83.74	83.83	91.26	92.11	92.22
29	101	82.80	83.39	89.85	96.49	98.16	96.89
30	90	82.27	83.07	86.91	86.81	88.30	88.72

Source: Own.

5 Final considerations

This paper considered the problem of variability in the daily counting of manufactured units in a production line and sent to quality control department of a food industry. It contributed illustrating how to identify probability distributions for times between arrivals and its parameters, helpful to construct models for Discrete Event Simulation Modeling or for applying on Queueing Theory models.

The statistical modeling of this data set presented, using Poisson regression models was very useful to indicate performance indicators of the industrial department (as a diagnostic device) in the company under study, using computational simulation. The discovering of important factors that affect the variability was very important in the decisions by the industrial managers in terms of reallocation of persons, due to small utilization average rate and in the programming of the statistical analyses (quality control), which implies in better planning of the production capacity, without affecting significantly the average lead-time of the quality verifications.

References

Ausín, M. C.; Wiper, M. P.; Lillo, R. E. Bayesian prediction of the transient behaviour and busy period in short- and long-tailed GI/G/1 queueing systems. Statistics Computational Statistics & Data Analysis, 52, p. 1615-1635, 2008.

Ausín, M. C.; Wiper, M. P.; Lillo, R. E. Bayesian estimation for the M/G/1 queue using a phase-type approximation. Journal of Statistical Planning and Inference, 118, p. 83-101, 2008.

Barossi-Filho, M.; Achcar, J.A.; Souza, R.M.(2010). Modelos de volatilidade estocástica em séries financeiras: uma aplicação para o IBOVESPA. *Economia Aplicada***,** Ribeirão Preto,14(1): 25-40.

Box, G.; Tiao,G. (1973). *Bayesian inference in statistical analysis*; New York: Addison- Wesley.

Cavalcante, C.A. V.; Almeida, A.T(2011) Modelo multicritério de apoio a decisão para o planejamento de manutenção preventiva utilizando PROMETHEE II em situações de incerteza. *Pesquisa Operacional*, Rio deJaneiro,25(2): 279-296.

Chib, S.; Greenberg, E. (1995). *Understanding the Metropolis-Hastings algorithm*. The American Statistician, 49, 327-335.

Droguett, E. L.; Mosleh, A (2011) Análise bayesiana da confiabilidade de produtos em desenvolvimento. *Gestão & Produção*, São Carlos,13(1): 57-69.

Ferreira, R.J. P.; Almeida Filho, A.T. ; Souza F. M.C. (2009) A decision model for portfolio selection. *Pesquisa Operacional,* Rio de Janeiro,29(2): 403-417.

Freitas, M.A. et al .(2010) Reliability assessment using degradation models: Bayesian and classical approaches. *Pesquisa Operacional*, Rio deJaneiro,30(1): 195-219.

Gamerman, D. (1997). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. London: Chapman and Hall.

Gelfand, A. E.; Smith, AFM (1990). Sampling-based approaches to calculating marginal distributions, *Journal of the American Statistical Association*, 85, 410, 398- 409.

Kalatzis,A. E. G.; Azzoni, C. R.; Achcar, J. A.(2006)Uma abordagem bayesiana para decisões de investimentos. *Pesquisa Operacional,* Rio de Janeiro,26(3): 585-604.

Motta, J.(1997) Decisões de preço em clima de incerteza: uma contribuição da análise Bayesiana. *Revista de Administração de Empresas*,São Paulo,37(2):31-46.

Moura, M.C. et al(2007) Avaliação Bayesiana da eficácia da manutenção via processo de renovação generalizado, *Pesquisa Operacional*, Rio de Janeiro,27(3): 569-589.

Paulino,C.D; Turkman,M.; Murteira,B.(2003). *Estatística Bayesiana*, Lisboa: Fundação Calouste Gulbenkian.

Quinino, R.C.; Bueno Neto, P. R.(1997). Avaliação bayesiana de inspetores no controle estatístico de atributos. *Gestão & Produção*, São Carlos, 4(3): 296-304.

Ramirez Pongo, R.M.; Bueno Neto, P.R.(1997). Uma metodologia bayesiana para estudos de confiabilidade na fase de projeto: aplicação em um produto eletrônico. *Gestão & Produção*, São Carlos ,4(3): 305-320

Sakurada, N.; MIYAKE, D. I. Aplicação de simuladores de eventos discretos no processo de modelagem de sistemas de operações de serviços. Gest. Prod. [online]. 16, 1, p. 25-43, 2009.

Shin, Y. W; Pearce, C. E. M. The BMAP/G/l vacation queue with queue-length dependent vacation schedule. Journal Austral. Math. Soc. Ser. B, 40, p. 207-221, 1998.

Spiegelhalter, D.J. , Best, N. G.; Carlin, B. P.; van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society*, series B, 64, 4, 583-639.

Spiegelhalter, D. J.; Thomas, A.; Best, N.G.; Lunn, D. (2003). *WinBugs: user manual*, version 1.4, MRC Biostatistics Unit, Cambridge, U.K

Wu, K.; Hui, K. The Determination and Indetermination of Service Times in Manufacturing Systems. IEEE Transactions on Semiconductor Manufacturing, 21, 1, p. 72-82, 2008.