Revista Produção On Line
Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

GRÁFICOS DE CONTROLE BASEADO NOS RESÍDUOS DO MODELO DE REGRESSÃO POISSON*

CONTROL CHART BASED ON THE POISSON REGRESSION MODELS RESIDUALS*

Suzana Russo
Dra. em Engenharia de Produção – UFSC
Universidade Federal de Sergipe - UFS
Núcleo de Graduação em Estatística - NES
Av. Marechal Rondon, s/n Jardim Rosa Elze
CEP 49100-000 São Cristóvão - SE
79-2105-6729 – suzana.ufs@hotmail.com

Maria Emilia Camargo
Dra. em Engenharia de Produção – UFSC
Universidade de Caxias do Sul – UCS
Depto de Ciências Sociais Aplicadas
Av. Dom Frei Cândido Maria Bampi 2800. CEP: 95200-000
kamargo@terra.com.br

Robert Wayne Samohyl
PhD. Em Engenharia Industrial - U.P.I.S.U
Universidade Federal de Santa Catarina
samohyl@eps.ufsc.br

RESUMO

Gráficos de controle baseado nos resíduos de Poisson têm sido têm sido úteis para monitorar o número de não conformidade em um processo industrial. O modelo de regressão de Poisson é o mais popular dos modelos lineares generalizados, o qual é usado para modelar dados de contagem. O modelo de regressão de Poisson tem uma suposição de que a variância é igual á media, mas nem sempre isso acontece, em muitas situações tem-se encontrado que a variância

Revista Produção On Line
Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

é maior do que a média, e este fenômeno é denominado como superdispersão. Os dados usados nesse estudo são números de não conformidades da seção de tecelagem da Indústria Têxtil Oeste Ltda. Observou-se que esses dados têm uma grande variabilidade e possuem superdispersão. Assim, foi preciso aplicar os modelos de Regressão Poisson antes da utilização das técnicas dos gráficos de controle.

Palavras-chave: Modelos de regressão de Poisson; Gráficos de controle, Não conformidade

ABSTRACT

The control chart based on the Poisson residuals has been useful for monitoring the number of nonconforming in an industrial process. The Poisson regression model is the most popular used by Generalized Linear Models, which is used to model the event count data. The Poisson regression model has an assumption that the variance is equal the mean, but it does not always happen, in many situations, it has found that the variance is greater than the mean, and this phenomenon is called as overdispersion. The data used in this study are the number of nonconforming at weaving section in the Têxtil Oeste Industry Ltda. It was observed that these data have a great variability and have overdispersion. Thus, it was necessary to apply the Poisson regression models before the use of the control charts techniques.

Keywords: Poisson Regression Models, Control Charts, Nonconforming

1. INTRODUCTION

In SPC (Statistical Process of Control) two forms of variation are identified. Common causes, that are sources of variation that are unavoidable, and special causes of variation that can be corrected or eliminated. A process that is subject only to common causes of variation is said to be in-control. When special causes of variation occur, the process is said to be out of control. The purpose of SPC is to determine when a process is going out of control, so the process may be corrected (MONTGOMERY, 1997).

The traditional methodology of SQC (Statistical Quality Control) is based on a fundamental supposition that the process of the data is independent statistically; however, the data not always are independent. When a process follows an adaptable model, or when the process is a deterministic function, the data will be autocorrelated.

According to McCullagh and Nelder (1989) the Poisson regression models is a kind of specific generalized linear models (GLM), and the maximum likelihood method is used to estimate the parameters of Poisson regression models. When there is a difference in the

ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

parameters, that is, the variance value is bigger than the mean value, one considers that there is overdispersion, the variance value is smaller than the mean value, suggest that there is underdispersion. Count data are analyzed through Poisson regression models that very often show overdispersion, so that the assumption of equality between average and variance is not valid.

Evidences of overdispersion or underdispersion suggest that there is not a very good fit for the Poisson regression model. One solution to this phenomenon is a random effects approach of finite mixture. In contrast to random effects models which treat overdispersion as a nuisance factor that complicates statistical inference, finite mixture models may provide additional insights about different sources of heterogeneity in the population (BÖCKENHOLT , 2000).

In this paper we present the methodology of Poisson regression models applied to real data sets in the number of nonconforming at weaving section in the Industry Têxtil Oeste Ltda in September, 2001, and analyze if the data have overdispersion.

2. POISSON REGRESSION MODEL

In accordance with McCullagh and Nelder (1989), the Poisson regression model is a specific kind of generalized linear model (GLM) it can be estimated using maximum-likelihood, with the following likelihood function

$$L = \prod_{i=1}^{n} \Pr(Z_i / \lambda_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}.\lambda^{Z}}{Z!} \tag{1}$$

and the following log-likelihood function $\log L = \sum (Z_i.\log(\lambda_i) - \lambda_i)$

The use of log link function assures that the fitted values of λ_i remain in the interval $[0,\infty)$. The Poisson regression model with a log link is sometimes called a loglinear model (McCULLAGH & NELDER, 1989 ; SCHAFER, 1997).

Transforming the log link function we obtain the following expression for the response variable:

$$r = e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k} \tag{2}$$

where $\beta \; x_i^T$ that is the linear function, x_i is the vector of covariates for the observation *i* and β is the unknown parameter (BÖCKENHOLT , 2000).

ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

Statistical of interest – After convergence (which it can be made through Newton-Raphson algorithm), it is necessary to examine the following statistics:

- $(X^T WX)^{-1}$, the estimated covariance matrix for $\hat{\beta}$;
- a log-likelihood function $l = \sum_{i=1}^{N} \{Z_i \log \lambda_i - \lambda_i\}$

For McCullagh and Nelder (1989) the deviance G^2 is the log-likelihood ratio statistic to compare the proposed model $\log \lambda_i = x_i^T \beta$, with the satured model. The satured model adapts a separated average for each Z_i, not considering the way the covariates are related. It easy to show that the likelihood $Z_i \log \lambda_i - \lambda_i$ it arrives at a maximum $\lambda_i = Z_i$, the deviance is

$$G^2 = 2\sum_{i=1}^{N} \left\{ Z_i \log \frac{Z_i}{\lambda_i} - (Z_i - \lambda_i) \right\} \tag{3}$$

Another way to check the best fit is the Pearson Chi-Square (χ^2) statistic, that compare observed distribution with the determined of the model (it makes us of the table of χ^2). The marginal adjustment of the Poisson regression model can then be determined by calculating the Pearson Chi-Square residuals: $r_i = \frac{Z_i - \lambda_i}{\sqrt{\lambda_i}}$, and the Pearson statistics for the best adjustment

$$\chi^2 = \sum_{i=1}^{N} r_i^2 \; . \tag{4}$$

Deviance and Pearson Chi-Square divided by the degrees of freedom are often used to detect overdispersion or underdispersion. For the Poisson distribution the mean and the variance are equal, which implies that the deviance and the Pearson statistic divided by the degrees of freedom should be approximately one.

The way to see the overdispersion is to suppose that var(Z_i)=$\lambda_i \, \sigma^2$ for some assumed value of σ^2>0. For σ^2=1 the model can be adjusted by Poisson, for other values of σ^2, it can be adjusted by the likelihood model. This represents to adjust for likelihood a equivalent negative binomial model (gamma-Poisson) (RUSSO, 2002; SCHAFER, 1997).

Revista Produção On Line
Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

3. EXAMPLE

The series used in this study number of nonconforming at weaving section in the Industry Têxtil Oeste Ltda, in September, 2001. The used method in the adjustment involves Poisson regression models. The daily number of nonconforming in the weaving section, 30 values, is shown in figure 1. Notice that the series have a great variability, and the data, apparently don't present the trend over time that was confirmed later.

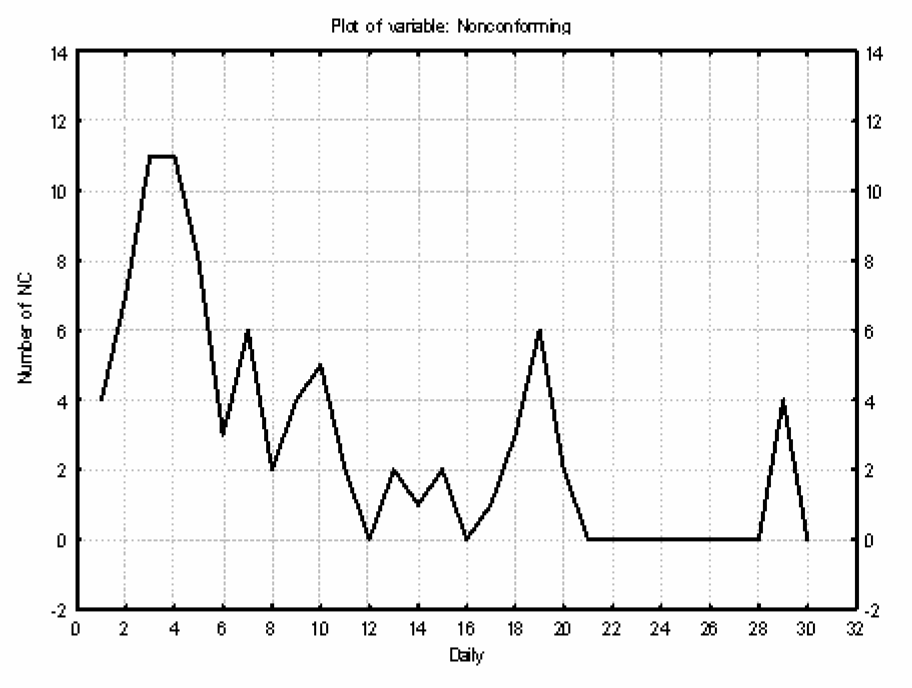

Figure 1 – Daily counting of the data

The number of nonconforming mean is 0,9508, and standard deviation is 1,16. Suppose that there is a tendency on the time, it is needed to have a polynomial of degree 4 to capture this effect. It was made the adjustment of the data through the Poisson regression models, where we found the following results:

Table 1 – Summary of estimative

	Coef	SE	Coef/SE	P val
X	1,036	0.179	5,788	0.000

Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

The algorithm of Newton-Raphson it converges in 5 iterations. The summary of analyzed model is shown in table 2.

Table 2 – Summary of analyzed model

Criterion	DF - Degrees of Freedom	Value	Value/DF
Scaled Deviance (G^2)	29	112,4591	3,8779
Scaled Pearson's (χ^2)	29	108,8571	3,7537
Loglikelihood		-81,3800	

The Pearson's statistics (χ^2) and the deviance (G^2) approach a χ^2 distribution, with the same degrees of freedom. Divide Pearson's statistics (χ^2) and the deviance (G^2) residuals by respective degrees of freedom are used to detect if there is overdispersion or underdispersion. For our data there is evidence of overdispersion; therefore the variance (1,3456) is larger than the mean (0,9508).

The large values of *Values/DF* indicate that the Poisson model doesn't fit well. We need to check the overdispersion. Analyzing the overdispersion we can find:

Table 3 – Summary of analyzed model

Criterion	DF - Degrees of Freedom	Value	Value/DF
Scaled Deviance (G^2)	29	29,0000	1,0000
Scaled Pearson's (χ^2)	29	28,0712	0,9679
Loglikelihood		-83,2264	

Table 3 indicates that the Poisson model is adjusted. In table 4, we find the summary of the parameter model.

Table 4 – Analysis of Parameter Estimates

Parameter	Estimate	Std Error	p val
Intercept	1,0290	0,2148	0,0000
Scale	1,9692	0,0000	

Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

According to Piegorsch (1998) graphic diagnoses and other operations about residuals help to indicate the discrepancy of the potential model.

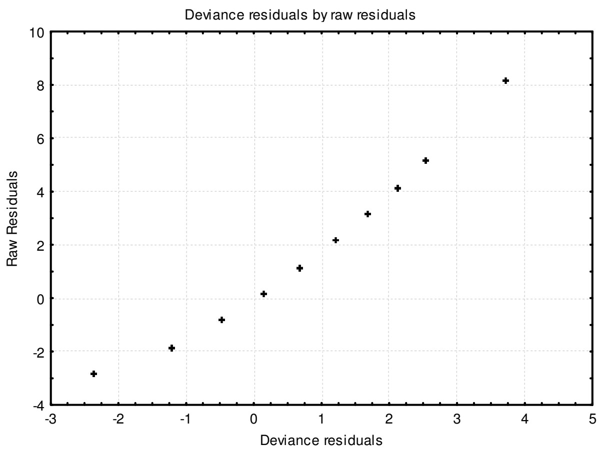

Figure 2 – Graph of the residuals of the deviance (G^2)

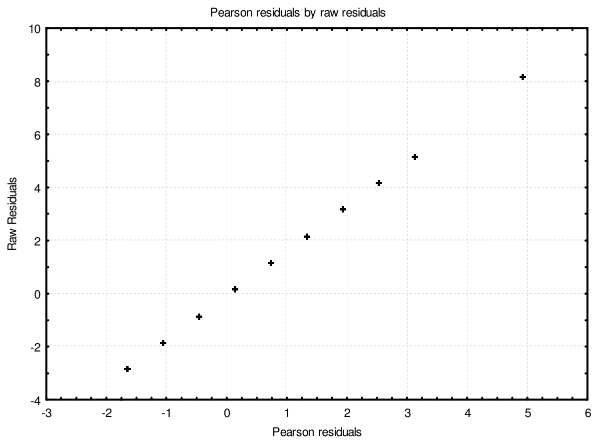

Figure 3 – Graph of Pearson's residuals

Figures 2 and 3 show that both types of residuals are equivalent for the observations what means that the residues of the Poisson regression model come relatively stable (PIEGORSCH, 1998).

Revista Produção On Line
Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

Application of the control chart

Now the productive process' behavior can be verified. Figures 4 and 5 show the control conditions for the observations in the C chart.

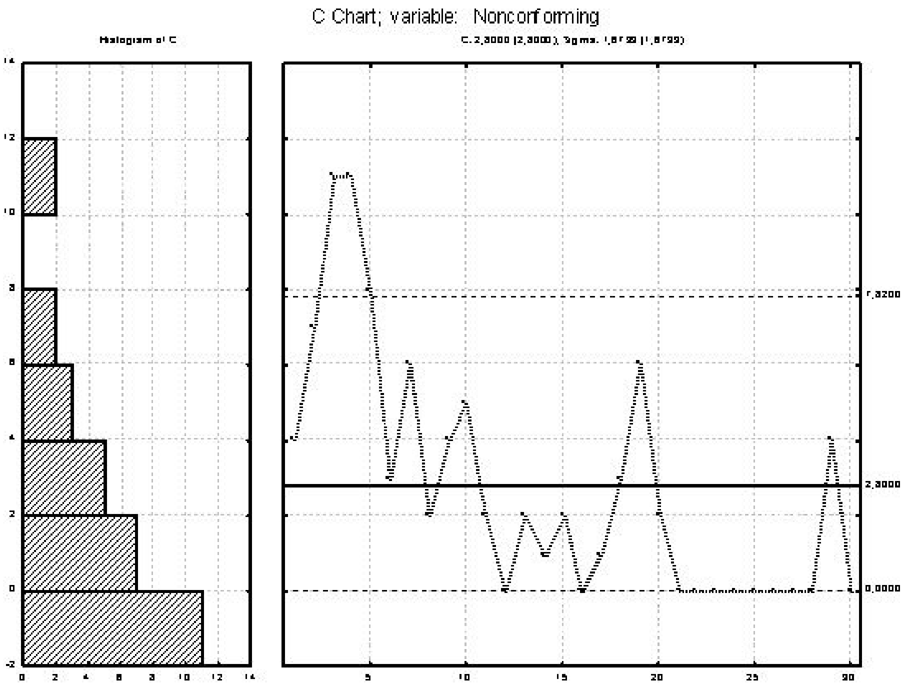

Figure 4 – C chart for real data.

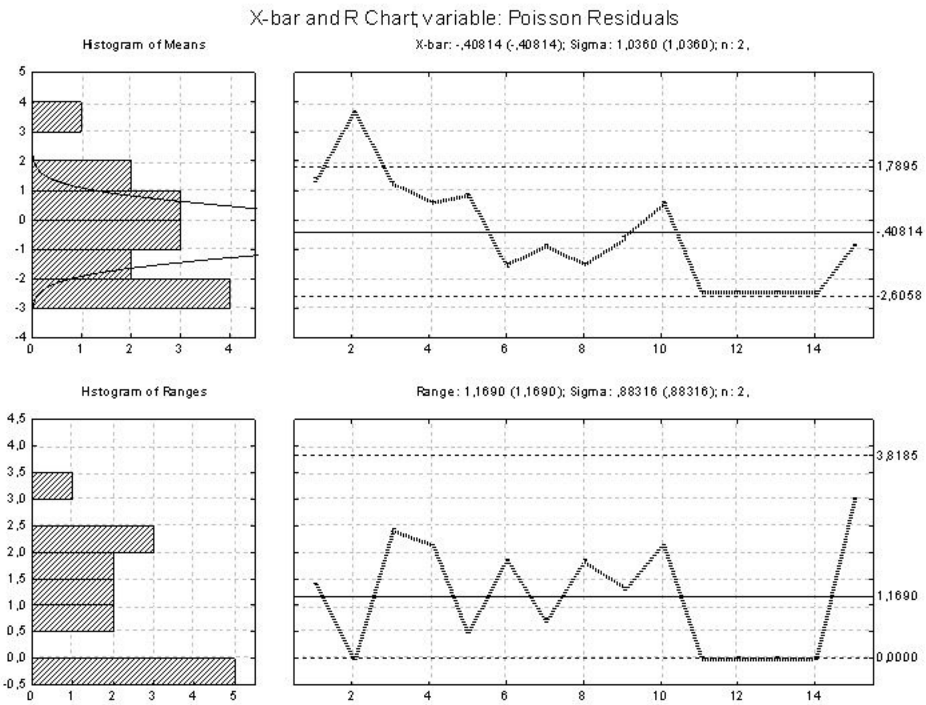

Figure 5 – $\overline{X}$ and R chart for transformed data.

Associação Brasileira de Engenharia de Produção - ABEPRO
Universidade Federal de Santa Catarina - UFSC
www.producaoonline.org.br
ISSN 1676 - 1901 / Vol. VIII/ Num. IV/ 2008

Figure 4 shows the C chart for real data and figure 5 shows the $\overline{X}$ and R chart for transformed data. Verifications revealed that the system wasn't adjusted and the C control chart was replotted with the found residuals of Poisson regression model to correct the problem. The problem was in the 3rd., 4th and 5th observations, which after having modeled for the Poisson regression model they were verified inside of the control limits, except the 2nd observation.

Acording to Wardell, Moskowitz and Plante (1994) it is entirely possible in traditional control charts, the points are out of the limits because of the systematic or the common causes and not because of occurrence of special causes.

4. CONCLUSION

In a monitored process, the variable must be independents, but if the intervals of time of monitoring are small, the data will be certainly autocorrelated. A chart of these data will frequently have a type in control, denoting a growth or decrease in the tendency. It means that these signs either they are really only false alarms out of the control, or the control limits are very conservative or even the data will have to be transformed because they are data autocorrelated.

The proposed methodology was shown satisfactory statistically; when data autocorrelations are studied a new learning perspective was generated on the productive process through of the information contained in the autocorrecion's structure, of the Poisson regression models, which were unknown for the classic model of monitoring. With this there was a growth of information for the correct decision and it can be detected that there was an improvement in the points of control exit.

As suggestion for further research, we recommend the use of the intervention analysis, in case the data have outliers; the use of non-parametric control charts and the use of other link functions.

REFERENCES

BÖCKENHOLT, U.*Mixed INAR(1)* Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. *Journal of Econometrics*. V.89 pp.317-338. 2000.

McCULLAGH, P.; NELDER, J. A. *Generalized Linear Models* .Second Edition. London: Chapman and Hall. 1989

MONTGOMERY, D. C. *Introduction to statistical quality.* 3ed. v.2, USA: John Wiley e Sons. 1997

PIEGORSCH, W. W. An introduction to binary response regression and associated trend analyses. *Journal of Quality Technology* . v.30 n.3 pp.269-281. 1998

RUSSO, S. L. *Gráficos de Controle para Variáveis Não-Conformes Autocorrelacionadas*. Brasil: Tese de Doutorado. Programa de Pós-graduação em Engenharia de Produção, UFSC, Florianópolis/SC. 2002

SCHAFER, J. L. *Analysis of incomplete multivariate data.* London. Chapman & Hall. 1997

WARDELL, D. G.; MOSKOWITZ, H.; PLANTE, R. D. Run-length distributions of special-cause control charts for correlated processes. *Technometrics*. v.36 n.1 pp. 3-17. 1994

Artigo recebido em 2006 e aprovado para publicação em 2008