Turkish Journal of Science & Technology
Volume 9(1), 89-96, 2014

Modeling of The Number of Divorce in Turkey Using The Generalized Poisson, Quasi-Poisson and Negative Binomial Regression

E. PAMUKCU[1], C. COLAK[2], N. HALISDEMIR[1]

[1]Department of Statistics, Firat University, 23119, Elazig, TURKEY
[2] Department of Biostatistics and Medical Informatics, Inonu University, Malatya, TURKEY
epamukcu@firat.edu.tr

(Received: 02.02.2013; Accepted: 20.09.2013)

Abstract

In this study, it has been aimed to model the numbers of divorce in Turkey between years 2001- 2009 using Generalized Poisson, Quasi-Poisson and Negative Binomial Regression methods. Data set of this study has been based on the data obtained from Turkish Statistical Institute (TUIK). Response variable-the annual rate of divorce- has been categorized into four groups with respect to the length of ex-married life of divorced couples. Explanatory variables have been designated as average age of the first marriage of men and women, the professional work life ratio of married women, the percentage of university graduates in both men and women. For Poisson models, overdispersion parameters have been detected respectively 32.413, 7.277, 16.158 and 26.361. Furthermore Pearson and G^2 statistics have revealed that Poisson models are not appropriate for data set. When Quasi Poisson regression was employed, it has been detected that residual deviances are rather close to Poisson residuals. Finally, Negative binomial regression has been conducted.
Overdispersion is a common phenomenon in Poisson modeling. In such data sets certain generalizations of Poisson regression and negative binomial regression modeling are used. In present study negative binomial regression has been detected as approved method.

Key Words: Generalized Poisson, Quasi-Poisson, Negative Binomial Regression, Overdispersion, Divorce

Türkiye'deki Boşanma Sayılarının Genelleştirilmiş Poisson, Quasi Poisson ve Negatif Binomiyal Regresyon Kullanılarak Modellenmesi

Özet

Bu çalışmada, 2001-2009 yılları arasında Türkiye'deki boşanma sayılarının Genelleştirilmiş Poisson, Quasi Poisson ve Negatif Binomiyal Regresyon metotlarına gore modellenmesi amaçlanmıştır. Çalışmanın veri seti, Türk İstatistik Kurumu (TÜİK)'den elde edilmiştir. Cevap değişkeni olan yıllık boşanma sayısı, boşanmış çiftlerin evil kalma sürelerine göre dört gruba ayrılmıştır. Çalışmanın veri seti Türkiye İstatistik Kurumu (TÜİK)'ndan elde edilen bilgiler ile oluşturulmuştur. Cevap değişkeni olan yıllık boşanma sayısı, boşanan çiftlerin evli kalma sürelerine göre dört gruba ayrılmıştır. Açıklayıcı değişkenler olarak, erkek ve kadınların ilk evlilik yaşı ortalamaları, evli kadının iş hayatına katılma oranı, erkek ve kadınlarda yüksek okul mezunu olma oranları ele alınmıştır. Poisson modelleri için aşırı yayılım parametresi sırasıyla 32.413, 7.277, 16.158 ve 26.361 olarak belirlenmiştir. Ayrıca Pearson ve G^2 istatistikleri de Poisson modellerinin veri seti için uygun olmadığını göstermiştir. Quasi Poisson uygulandığında ise artıkların dağılımı Poisson modellerine çok yakın çıkmıştır. Sonuç olarak Negatif Binomiyal Regresyon kullanılmıştır. Aşırıyayılım, Poisson modellemesinde yaygın bir fenomendir. Bu gibi veri setlerinde Poisson Regresyonun çeşitli genelleştirmeleri ve Negatif Binomiyal Regresyon kullanılır. Bu çalışmada Negatif Binomiyal Regresyonun uygun olduğuna karar verilmiştir.

Anahtar kelimeler: Genelleştirilmiş Poisson, Quasi Poisson, Negatif Binomiyal Regresyon, Aşırıyayılım, Boşanma

Bu çalışmanın özeti, sözlü sunum olarak 7. Uluslararası İstatistik Kongresi'nde tebliğ edilmiştir

1. Introduction

One of the main problems of modern age is the striking change that the family, marital relations in particular, has gone through. Dissolutions in traditional family composition, differences amongst generations, noticeable rise in divorce ratio of and spread of cohabitation have become hard-and-fast facts of present day. These modifications are mostly related to socially-based personal motives the foremost of which are financial mishaps in family, the rise in financial independence of women and changing role of women in family [1].

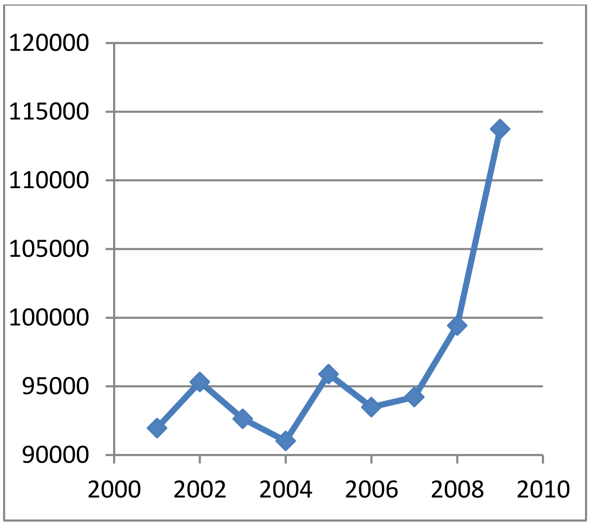

Figure 1: The number of divorces in Turkey with respect to years

In recent years, the birth rates have dramatically fallen throughout the whole world, in developed countries particularly, in addition to the rise in marriage age and first child birth as well as the shrinkage in family size. The transformation of traditional family structure into nuclear family has introduced the chain of changing relationships within family and the consequential status of kids and "single-parent" families which are particularly widespread in western societies. The rise and spread of divorcements leads to an increase in single-parent families. The change in male-dominated family concept and educational-professional rise of women which offers financial independence to females are considered to be the most significant factors affecting family structure.

Another element accelerating the divorce trend is that in almost all western states it has become rather easy to get a divorce. In addition to that, the destruction of taboos against divorce, couples' search for happiness, age of first marriage, the length of ex-married life, socio-economic factors and education are also influential elements which are all effective in weakening patriarchal structure dominating families [2].

In current study the attempt has been to evaluate within this context the factors accelerating the latest rise in divorce ratios in Turkey; hence the average age of the first marriage in both men and women, the professional work life ratio of married woman, the ratio of possessing college diploma in both men and women have been taken as explanatory variables.

2. Material and Method

2.1. Generalized Poisson Regression (GPR)

The most widely used regression model for count data sets is Poisson regression model which includes variables into the model through log-link function. The most distinctive feature of Poisson model is that they have equi-dispersion; nonetheless in applications, data sets generally possess a variance that exceeds the mean. In such cases excess variability phenomenon is defined as overdispersion [3,4]. When there is overdispersion in data set, generalized Poisson distribution is such;

$$f(y_i;\theta_i,k) = \frac{\theta_i(\theta_i + ky_i)^{y_i-1}e^{-\theta_i-ky_i}}{y_i!}$$

$$y_i = 0,1,2 \ldots \qquad (1)$$

Here, $\theta_i > 0$ and $\max\left(-1, \frac{-\theta_i}{4}\right) < k < 1$. Also expected outcome and variance of generalized Poisson distribution are as;

$$\mu_i = E(Y_i) = \frac{\theta_i}{1-k},$$

$$Var(Y_i) = \frac{\theta_i}{(1-k)^3} = \frac{1}{(1-k)^2}$$

$$E(Y_i) = \phi E(Y_i) \quad (2)$$

Specifically $\phi = \frac{1}{(1-k)^2}$ term plays the role of a dispersion factor. It is evident that for $k = 0$, generalized Poisson dispersion is θ_i parametered general Poisson distribution. When $k < 0$ there is underdispersion whereas if $k > 0$ there is overdispersion [5]. Overdispersion problem leads to serious underestimation of standard error and misleading inference for the regression parameters. Consequently a number of estimation methods have been proposed for modeling overdispersed data. These are the models that include quasi-Poisson or quasi-binomial regression model and negative binomial distribution. Parameter estimations of these models are similar to simple Poisson approach yet confidence intervals are larger [6]. Consequently the models are to provide different outcomes with respect to the meaningfulness of coefficients.

2.2. Quasi-Poisson and Negative Binomial Regression Models (QP and NBR)

These two methods which generally provide similar results can demonstrate striking divergences in estimating the effects of variables as well. The primary reason accounting for this difference is that: In NBR, variance is quadratic function of the mean whereas in QP model variance is linear function of the mean. Therefore, in QP model, for a Y random variable the equation is:

$$E(Y) = \mu$$
$$Var(Y) = \phi\mu \quad (3)$$

Hereby $E(Y), Y$ is the expected value of the latter ***(Y)*** and also known as the mean of distribution. $Var(Y), Y$ is the variance of the latter ***(Y)***. $\mu > 0$ and $\phi > 1$. For $\phi = 1$ it is obvious that Y is Poisson distributed. In equation (3) ϕ is overdispersion parameter. For NBR model,

$$E(Y) = \mu$$
$$Var(Y) = \mu + a\mu^2 \quad (4)$$

and overdispersion parameter is $\phi = (1 + a\mu)$. Contrary to QP model, overdispersion parameter depends on μ.

One of the main problems related to selecting these models is this: What is the effect of the use of NBR and QP in estimating β regression coefficients? Indeed the different relationship between means/variance ratio proves that regression coefficients of NBR model and QP models can also be not the same as well. Because fitting these models NBR and QP use weighted least-squares method and these weights are inversely proportion to variance. In that case NBR and QP shall weight the observations differently. For QP, weights are directly proportional to the mean while NBR weights have a concave relationship to the mean [7].

In present study all three methods have been applied to the data set with an overdispersion problem and containing the divorce numbers between 2001-2009 in Turkey in addition to relevant explanatory variables. Hence it has been aimed to analyze relative validity of the estimations.

2.3. Pearson Statistics

The standard goodness-of-fit measurement method for a μ averaged and ω variance dependent variable Y is Pearson statistics which is formulated as;

$$P = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\widehat{\omega_i}} \quad (5)$$

This value is used to determine whether the dispersion of series is excessive. Here the values $\hat{\mu}$ and $\widehat{\omega_i}$ are the predicted values of μ_i and ω_i . Calculated P value is compared with degree of freedom $n - k$designated for $\hat{\mu}$. When this formula is applied for Poisson regression, $\hat{\mu} = \hat{\omega}$ is the outcome and the final result shall be;

$$P_p = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\hat{\mu}} \quad (6)$$

Calculated P_p value shall similarly be compared with $n - k$.

2.4. Deviation Statistics

Another technique employed in measuring goodness-of-fit is deviation statistics. This statistics value is also known as “G square statistics”. G square statistics is expressed such;

$$G^2 = \sum_{i=1}^{n} Y_i \ln(\frac{Y_i}{\mu}) \quad (7)$$

If this statistics value approaches to 0 it indicates an increase in fitting of model. If this statistics value equals to exactly 0, it is reasonable to assert that fitting of model is perfect [8].

3. Application

In the application four separate models have been created according to the lengths of ex-married life of divorced couples. Data set of current study has been based on information obtained from Turkish Statistics Institute (TUIK) and in statistical analyses R 2.11.1 software has been benefited. As indicated in Table-1, data sets employed for each single model are such: Y_1: the number of divorced couples who remained married for 0-5 years, Y_2: the number of divorced couples who remained married for 6-10 years, Y_3: the number of divorced couples who remained married for 11-15 years, Y_4: the number of divorced couples who remained married for 16+ years, X_1: Average age of marriage in men, X_2: Average age of first marriage in women, X_3: Work life participation ratio in married women, X_4: The ratio of possessing college diploma in men, X_5: The ratio of possessing college diploma in women.

Table 1. Data set used in the research

Year	Number of divorces with respect to the length of ex-marriage				$X_{1(\%)}$	$X_{2(\%)}$	$X_{3(\%)}$	$X_{4(\%)}$	$X_{5(\%)}$
	0-5 (Y_1)	6-10 (Y_2)	11-15 (Y_3)	16+ (Y_4)					
2001	39065	20089	13125	19685	25.5	22.2	25.9	13.12	11.38
2002	40190	20726	13545	20862	25.9	22.7	26.4	13.75	12.17
2003	39500	19690	12587	20860	25.9	22.7	25.3	15.73	13.53
2004	39386	19129	11833	20674	26	22.8	21.6	16.62	13.93
2005	40725	20454	12900	21816	26.1	22.8	21.3	18.03	15.1
2006	39817	20387	12660	20625	26.1	22.8	21.5	18.85	17.41
2007	39420	20870	12757	21172	26.1	22.8	21.6	21.56	18.66
2008	41228	21335	13863	22997	26.2	22.9	22.4	22.37	19.69
2009	45803	23879	16628	27426	26.3	23	24.3	29.4	25.92

Table 2. Descriptive Statistics

Variables	n	Minimum	Maximum	Mean	SD
Y_1	9	39065.00	45803.00	40570.44	2082.27
Y_2	9	19129.00	23879.00	20728.77	1348.67
Y_3	9	11833.00	16628.00	13322.00	1369.47
Y_4	9	19685.00	27426.00	21790.77	2301.18
X_1	9	25.50	26.30	26.01	0.23
X_2	9	22.20	23.00	22.74	0.22
X_3	9	21.30	26.40	23.36	2.09
X_4	9	13.12	29.40	18.82	5.06
X_5	9	11.38	25.92	16.42	4.56

Descriptive statistics pertaining to response variables and explanatory variables are as shown in Table-2. Accordingly for all the variables the means are 40570±2082 for Y_1, 20729±1349 for

Y_2, 13322±1369 for Y_3, 21790±2301 for Y_4, 26.01±0.23 for X_1, 22.74±0.22 for X_2, 23.37±2.09 for X_3, 18,83±5.07 for X_4, 16.42±4.56 for X_5. To detect if there is a difference in number of divorces with respect to the length of ex-marriage, variance analysis has been implemented on response variables and in the end the difference in number of divorces with respect to the length of ex-marriage has been found statistically meaningful ($p<0.001$).

According to Bonferroni multiple comparison test result, no statistically meaningful difference has been detected between groups Y_2 and Y_4 yet amongst the rest of all groups a statistically meaningful difference has been identified. Goodness-of-fit measurements pertaining to the methods applied to four separate models are as demonstrated in Table-3.

Table 3. GPR and NBR goodness-of-fit measurements for four models

	Generalized Poisson Regression				Negative Binomial Regression			
	Df	Calculation value	ϕ	AIC	Df	Calculation value	ϕ	AIC
Pearson[1]	3	97.238	32.412	221.72	3	8.915	2.972	156.46
Deviation[1]	3	32.564	10.854		3	2.998	0.999	
Pearson[2]	3	21.832	7.277	139.88	3	8.999	2.999	137.06
Deviation[2]	3	7.302	2.434		3	3.016	1.005	
Pearson[3]	3	48.474	16.158	162.72	3	8.934	2.978	140.34
Deviation[3]	3	16.248	5.416		3	3.010	1.003	
Pearson[4]	3	79.084	26.361	198.18	3	8.839	2.946	148.97
Deviation[4]	3	26.590	8.863		3	2.996	0.998	

Pearson[a], Deviation[a]: Values calculated for a. Model.; Df: Degree of freedom;
ϕ:Overdispersion parameter; AIC: Akaike information criteria

In the beginning Generalized Poisson Regression has been applied to models. Akaike Information Criteria (AIC) values of models have been obtained respectively 221.72, 139.88, 162.72, and 198.18. Goodness-of-fit of all new models has been analyzed via Pearson and G^2 statistics. Since in each single model $p<0.001$ has been the outcome for Pearson and G^2 statistics it has been agreed that Poisson models are not fit for data set. Assuming the existence of overdispersion in data set, dispersion parameters have been calculated respectively for each model.

In the researches, model fit is based on the measurements of deviation and Pearson chi-square fit. Both measurements indicate approximate chi-square distribution. Dispersion parameter values are obtained through dividing calculation value into degree of freedom. Dispersion parameter values concerning deviation fit measurement in Poisson regression have been found respectively for each single model as 10.854, 2.434, 5.416, 8.863 and dispersion parameter values of Pearson chi -square fit measurement have been found respectively for each single model as 32.413, 7.277, 16.158 and 26.361.

Since all obtained dispersion values were larger than 1, it has been agreed to apply alternative models.

Quasi Poisson regression makes estimations based on above-calculated dispersion parameters as well. When quasi Poisson was applied to all models, AIC could not be calculated due to the lack of sufficient data, residual deviations have been rather close to the residuals of Poisson models and therefore the superiority of models over one another could not be detected. When Negative Binomial Regression was applied dispersion parameter values of Pearson chi-square fit measurement have been obtained respectively for each model as 2.972, 2.999, 2.978, and 2.946. The comparison with Poisson models revealed that dispersion parameters approach to 1. Besides the fact that deviation fit measurement values were close to 0 for each model respectively 0.999, 1.005, 1.003, 0.998 indicated that compared with Poisson models, fitting of model is better. Obtained parameter estimations are as given below

Table 4: Negative binomial regression parameter estimations

NB models	Variables	Beta	S.E.	P	AIC
	Constant	6.797	2.448	**	
Model1	X_3	0.015	0.005	**	156.46
	Constant	3.422	1.636	*	
	X_1	0.724	0.171	***	
Model2	X_2	-0.569	0.133	***	137.06
	X_3	0.023	0.003	***	
	X_1	0.979	0.319	**	
Model3	X_2	-0.775	0.249	**	140.34
	X_3	0.042	0.006	***	
	X_3	0.025	0.006	***	
Model4	X_4	0.048	0.015	**	148.97
	X_5	-0.040	0.017	*	
p codes	: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

AIC:Akaike information criteria; SE:Standart error

Model-1: $\log(\mu)= 6.797+0.015*X_1$,
Model-2:$\log(\mu)=3.422+0.724*X_1-0.569*X_2+0.023*X_3$,
Model-3:$\log(\mu)=0.979*X_1-0.775*X_2+0.042*X_3$,
Model-4:$\log(\mu)=0.025*X_3+0.048*X_4-0.04*X_5$

4. Discussion and Conclusion

Divorce is a complex phenomenon that can be analyzed from a different set of perspectives. It has been witnessed that sociologic researches related to divorce basically focus on structural and vital determiners and it has been noted that these determinants change with respect to the social class, genders of people and age of first marriage. Psychological researches on the other hand stress upon inter-family communication, personal traits in addition to chronic negative effects such as conflicts within family or anti-social conduct.

Within that context, as the studies relevant of the causes of divorce are analyzed it has been detected that they are mostly retrospective studies that cover socio-demographic features which emerge upon analyzing divorcement cases or the divorced couples who consult to a crisis center [9,10]. There are very few number of studies that include divorcement-based statistical analyses. Another illustrative example can be shown as the study of Kei Sakota who analyzed divorcement ratios between 1964-2006 in Japan via time series[11]. In a research conducted by Maxin and Berec (2010), it has been established through a demographic model that the rise in the number of single individuals in society (not married ever or divorced after a marriage) positively affects the rate in divorcements [12]. Furthermore as the similar studies on divorce are analyzed it has been observed that not the causes but rather the economic and social consequences of divorcement have been explored.

It is likely to assume that the difficulty in expressing a social phenomenon like divorce through a mathematical model may be the reason accounting for the limited number of relevant studies. It is without question that a deeper look into divorce cases reveals the fact that the most common causes of divorcement are severe conflict, domestic violence, financial problems and so on. The question is what might be the hidden causes lying beneath these explicit reasons? What might be roots of precedent problems that put the individuals on the threshold of divorce?

In recent years, the birth rates have dramatically fallen throughout the whole world, in developed countries particularly, in addition to the rise in age of first marriage and first child birth as well as the shrinkage in family size. The transformation of traditional family structure into nuclear family has introduced the chain of changing relationships within family and the consequential status of kids and "single-parent" families which are particularly widespread in western societies. The rise and spread of divorcements leads to the increase in single-parent families. The change in male-dominated family concept and educational-professional rise of women which offers financial independence to females are considered to be the most significant factors affecting family structure. The destruction of taboos against divorce, couples' search for happiness, age of first marriage, the length of ex-married life, socio-economic factors and education are also influential elements which are all effective in weakening patriarchal structure dominating families [2].

In current study the attempt has been to evaluate within this context the factors accelerating the latest rise in divorce ratios in Turkey; hence the average age of the first marriage in both men and women, the professional work life ratio of married woman, the ratio of possessing college diploma in both men and women have been taken as explanatory variables.

As the numbers of divorces in Turkey between years 2001-2009 were analyzed, the problem of overdispersion surfaced and amongst the regression models applied to solve this problem negative binomial regression model has been deemed to be the most appropriate one. All explanatory variables used in this research have been found to be meaningful in different combinations depending on response variable. Particularly in men, the rise in average age of first marriage is positively influential in the number of divorces which may be attributed to the reason that males who remain and live single for long periods of time face hardships in adapting to marital life and its consequential responsibilities. In women on the other hand the average age of first marriage affects the number of divorces negative exponentially; in other words in a positive direction yet insignificantly. In this instance the effective factor might be the pressure society and people exert on women against marrying at old ages which inevitably motivates women to sustain this long-sought marital life. Once married woman enters into laborforce she no longer values marriage as her insurance hence it becomes easier for her to annul a marriage. Consequently the shift in laborforce participation ratios of women has positively affected the models. As regards men, the rise in college diploma and the upcoming interval to embark on work life inevitably increase the age of first marriage. Therefore above-mentioned situation that is related to the age of first marriage in men holds true in this case as well. The same finding shall be valid in cases when women are college graduates. In other words, a woman who delays marriage to possess a college diploma and profession meets family life- which is stimulated through social pressure- late so she desires to sustain this delayed status of being a married woman. Occasionally this may affect the number of divorces negative exponentially.

To sum up, in present research where divorce phenomenon is analyzed with respect to certain risk factors that may be effective in inter-family relations, overdispersion has surfaced as a problem and in solving this problem some of the proposed alternative regression methods have been employed. NBR model amongst these methods can be suggested to be used in overdispersed data. Additionally due to the restricted number of studies employing advanced statistical methods related to divorce causes, the modern risk factors which may lead to divorcement should be evaluated under multi-centered and larger samples. All explanatory variables used in this research have been found to be meaningful in different combinations depending on response variable. The meaningfulness of variables can be evaluated with further sociological investigation.

5. References

1. Süleymanov, A., (2010). Family and marital relations in modern Turkish societies. Journal of Politics Conferences. p.: 198-216.
2. Kurtulmuş, S., (1998). The effect of changes in family models and demographic structure on

family policies. III. Family Council Papers. Ankara
3. Long, J.S., (1997). Regression Models for Categorical and Limited Dependent Variables. Sage Publisher. USA, p: 217-249
4. Dean, C.B., (1992). Testing for overdispersion in Poisson and Binomial Regression Models. JASA, 87(418).
5. Yang, Z., Hardin, J.W. and Addy, C. L., (2009). A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. Journal of statistical planning and inference. 139: 1514-1521
6. Logan, M., (2010). Biostatistical Design and Analysis Using R. Willey-Blackwell.
7. Hoef, J.M.V., Boveng, P.L., (2007). Quasi-Poisson vs negative binomial regression, How should we model overdispersed count data?. Ecology. 88(11): 2766-2772.
8. Deniz, Ö., (2005). Poisson Regression Analysis, Istanbul University of Commerce Journal of Sciences, 4(7):59-72
9. Demirci, Ş., Günaydın, İ.G., Doğan, K.H., Aynacı, Y., (2005). Retrospective analysis of the divorces in Konya, Journal of Forensic Science, 19(11): 22-28
10. Uçan, Ö., (2007). Retrospective analysis of the women consulting a crisis center in divorcement stage. Clinical Psychiatry , 10: 38-45
11. Sakata, K., McKenzei, C.R., (2009). The impact of divorce precedents on the Japanese divorce rate. Mathematics And Computers In Simulation (79):2917-2926
12. Maxin, D., Berec, L., (2010). A two-sex demographic model with single dependent rate, Journal of Theoretical Biology, (265): 647-656