Sample Size Determination For Follow-up Studies of Exponentiated Exponential Distribution

By

Saida M. Saleh

Department of Statistics - Faculty of Commerce

Al – Azhar University

Sample Size Determination For Follow-up Studies of Exponentiated Exponential Distribution

By

Saida M. Saleh

Department of Statistics - Faculty of Commerce

Al – Azhar University

Abstract

The objective of this study is to identify the appropriate sample size for specific follow-up. The hazard function of the exponentiated exponential distribution (EE) with shape parameter γ and scale parameter λ with covariates having coefficient B is assumed. A test of the null hypothesis $H_0 : B = B_0$ versus the alternative hypothesis $H_1 : B = B_1 < B_0$ at a specified level of significance (α_0) and a specified power ($1-\beta_0$) is considered. The hazard function and its reversed function can be used to compute the Fisher information matrix of the unknown parameters A general sample size formula for more than two dose groups for testing H_0 versus H_1 at level of significance (α_0) and with power ($1-\beta_0$) is derived. Applications involving three parameters, namely the shape, scale, and the coefficient of the covariates variable are developed. A numerical example will be carried out to illustrate the theatrical results using Mathcad (2001).

It is clear that the sample size increases when the required power increases or when the level of significance decreases, its increases when the length of the follow-up periods increases. The scale parameter λ increases as the sample size decreases.

Key words: the Follow-up studies; Exponentiated exponential distribution; Proportional Hazard models; Fisher information matrix, Sample Size.

1. Introduction

Follow-up studies play an important role not only in the investigations of the causes of chronic diseases and of the effects of preventive measures or treatments, but also in the survival analysis of industrial products or projects. The study starts by determining the sample size of individuals (patients or units of a certain product) to be

followed during the follow-up period. This follow-up period (or the period of observations) lasts till one of the following three events occurs; the individual drops out of the program study, the individual is lost (or dies) during the study period or the study is completed [Feigl and Zelen (1965), George and Desu (1974), , Taulbee and Symons (1983), and Kung (1993).]

Many investigators considered the problem of estimating sample size. Among those who have studied estimation of sample size Johnson(1962), Harman(1967), Guenther(1970),Gokhale(1972),George, and Desu, (1974), Narula, and Li, (1975), ,Ashour and Shalaby (1983), Ashour et.al.(1996), Abd-Elfattah and Bakoban(2003),.

The particular problem of interest here is sample size determination for follow up studies of exponentiated exponential distribution using approach of Saleh,(2002).

This article can be organized as follows. In section 2 the exponentiated exponential distribution is introduced. Section 3 presents the exponentiated exponential proportional hazard model. In section 4 the Fisher information matrix, the approximate asymptotic variances and covariances matrix are investigated. Section 5 explains the tests of the hypotheses about B. For illustrating the relationship between sample size and the length of the studies a numerical example will be carried out in section 6. In section 7 discussion and recommendations presents.

2. The Exponentiated Exponential Distribution.

The exponentiated exponential distribution (EE), also known as generalized exponential distribution has been studied quite extensively by Gupta and Kundu (1999, 2001a, 2001b, 2002,) and Nadarajah and kotz, (2006). It is observed that the EE exponential distribution can be considered for situations where a skewed distribution for a non-negative random variable is needed. Also, it is observed that it can be used quite effectively to analyze lifetime data, particularly in presence of censoring. It can be used as a possible alternative to the two-parameter Weibull and gamma distributions in many situations. One major disadvantage of the gamma distribution is that the distribution function cannot be expressed in a closed form if the shape parameter is not an integer, while the EE distribution function has a closed form. In addition, when the shape parameter of the Weibull and EE distributions are greater than one, the hazard functions of both distributions are increasing. However, in the case of the EE distribution it increases from a zero to a finite number, whereas in the case of the Weibull distribution it increases from zero to infinity. Therefore, if it

is known that the data are from a regular maintenance environment, it is better to fit the generalized exponential distribution than the Weibull distribution.

The EE distribution has the following cumulative distribution function

$$F(t;\gamma,\lambda) = (1-e^{-\lambda t})^{\gamma} \quad , \gamma,\lambda > 0 \qquad (2.1)$$

The corresponding density function is:

$$f(t;\gamma,\lambda) = \gamma\lambda(1-e^{-\lambda t})^{\gamma-1} e^{-\lambda t} \quad , \gamma,\lambda > 0 \qquad (2.2)$$

for $t > 0$ and 0 otherwise. Here γ is the shape and λ is the scale parameters. If the shape parameter $\gamma = 1$, then the EE distribution coincides with the exponential distribution with a scale parameter λ. The EE distribution function is a strictly decreasing function if $\gamma \leq 1$, whereas if $\gamma > 1$, it is a unimodal skewed density function.

The survival and hazard functions of the EE are given respectively by:

$$S(t;\gamma,\lambda) = 1-(1-e^{-\lambda t})^{\gamma}, \qquad (2.3)$$

and,

$$h(t;\gamma,\lambda) = \frac{\gamma\lambda(1-e^{-\lambda t})^{\gamma-1} e^{-\lambda t}}{1-(1-e^{-\lambda t})^{\gamma}}. \qquad (2.4)$$

The reversed hazard function becomes quite popular in the recent time. The reversed hazard function for the EE distribution is

$$r(t;\gamma,\lambda) = \frac{f(t;\gamma,\lambda)}{S(t;\gamma,\lambda)} = \frac{\gamma\lambda e^{-\lambda t}}{1-e^{-\lambda t}} \qquad (2.5)$$

It is observed that for all values of γ, the reversed hazard function is a decreasing function of t. Nanda and Gupta (2003) obtained several other properties of the reversed hazard function of the generalized exponential distribution. The hazard function and the reversed function can be used to compute the Fisher information matrix of the unknown parameters,(see for example Efron and Johnstone (1990)). For the EE distribution $r(t;\gamma,\lambda)$ is in a convenient form and it can easily be used to compute Fisher information matrix (see Gupta, and Kundu (2005)).

It can be noted that the EE(γ,λ) distribution has an increasing or decreasing hazard function if $\gamma > 1$ or $\gamma < 1$ respectively and for $\gamma = 1$ the hazard function is constant.

According to Gupta and Kundu (2005) reversed hazard function of the EE distribution will be used in this study.

3. The Exponentiated Exponential Proportional Hazard Model.

The exponentiated proportional hazard model (EEP) was suggested firstly by Cox (1972). He introduced the common form of the proportional hazard model and a suitable method for analyzing such a model if the baseline hazard function is considered as a nuisance parameter and main interest being in the regression parameters. The general form is:

$$h(t|Z)=h_0(t)C(ZB) \qquad (3.1)$$

where

$h(t|Z)$ is the conditional risk of experiencing the event at t for individuals having covariate vector Z consisting of b elements. $h_0(t)$ is the underlying hazard rate, i.e. the baseline function that depends on time and having covariate vector Z= 0 (i.e. under standard conditions). If one assume a particular form for $h_0(t)$,a fully parametric proportional hazard model obtained B is the $b\times 1$ column vector of unknown parameters through which the hazard depends on the covariates Z is a $1\times b$ row vector of covariates associated with an observation whose lifetime is t. $C(ZB)$ is a linking function between the hazard rate and the explanatory variables.

By the proportionality assumption model (3.1) can take the form:

$$h(t|Z)=h_0(t)e^{ZB}, \qquad (3.2)$$

with $h_0(t)$ following a specific parametric distribution.

Then, in view of equation (3.2),

$$r(t_i|Z_j)=h_0(t)e^{ZB}, \qquad (3.4)$$

From equation (2.4)

$$r(t;\gamma,\lambda|Z_j)=\frac{\gamma\lambda e^{-\lambda t}}{1-e^{-\lambda t_i}}e^{BZ_j} \qquad (3.5)$$

Assume that the covariate Z is discretely distributed with possible values Z_j where j= 1,.., J. let $S_j(t_i)= S_j(t_i \mid Z_j)$ then

$$S(t;\gamma,\lambda|Z_j)=\exp\left(-\int_0^{t_i} r(x|z_j)\,dx\right)$$

$$= \exp\left(- \int_0^{t_i} \frac{\gamma \lambda e^{-\lambda t}}{1 - e^{-\lambda t_i}} e^{B Z_j} dt \right)$$

$$= \exp\left(- \gamma e^{B Z_j} \ln(1 - e^{-\lambda t}) \right)$$

$$= (1 - e^{-\lambda t})^{-\gamma e^{B Z_j}} \qquad (3.6)$$

4. The Fisher information matrix:

The likelihood function summarizes the likelihood of all observation in the sample (see Lawless1982). If the observed individual; died at time t, then

$$f(t_i) = h(t_i)\, S(t_i) \qquad (4.1\ )$$

Suppose that there are n individuals under study, the associated lifetime of the individual is T_i and a fixed censoring time T_i. The T_i are assumed to be survival function in equation (3.6). The exact lifetime T_i of an individual will be observed only if $T_i \leq T_c$ (i.e. fixed time). The data from such an experiment can be represented by the n pairs of random variables (t_i, δ_i) where $t_i = \min(T_i, T_c)$, $\delta_i = 1$ if $t_{ij} \leq T_c$ and $\delta_i = 0$ if $T_{ij} > T_c$. Let Z be the covariate variable with level Z_j, $j = 1,\ldots,J$, the unites of a sample of size n_j are given dose j with level Z_j, $j = 1,\ldots,J$ and

$$n = \sum_{j=1}^{j} n_j$$

Let T_{ij} be the observed lifetime for individual i receiving dose j.

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad T_{ij} \leq T_c \\ 0 & \text{if} \quad T_{ij} > T_c \end{cases} \qquad (4.2)$$

The likelihood function used (4.1) has the following form

$$f(t; \gamma, \lambda | Z_j) = r(t_{ij} | T_c, Z_j)^{\delta_i}\, S(t_i | T_c, Z_j)^{1-\delta_i} \qquad (4.3)$$

$$L(\gamma, \lambda, B, T_c | t_{ij}, Z_j, i = 1,\ldots,n_j, j = 1,\ldots,J) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left[r(t_{ij} | T_c, Z_j)\, S(t_{ij} | T_c, Z_j) \right]^{\delta_i} \left[S(t_i | T_c, Z_j) \right]^{1-\delta_i}$$

$$L(\gamma, \lambda, B, T_c | t_{ij}, Z_j, i = 1,\ldots,n_j, j = 1,\ldots,J) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} r(t_{ij} | T_c, Z_j)^{\delta_i}\, S(t_i | T_c, Z_j)$$

The logarithm of the likelihood function is

$$Log\ L = \sum_{j=1}^{J}\sum_{i=1}^{n_j} \delta_{ij} \log r(t_{ij}|T_c, Z_j)' + \log S(t_i|T_c, Z_j)' \qquad (4.4)$$

Assume that the shape parameter $\gamma = \gamma_0$. The approach here is to use models $r(t_{ij}|T_c, Z_j)$ and $S(t_{ij}|T_c, Z_j)$

$$Log\ L = \sum_{j=1}^{J}\sum_{i=1}^{n_j} \delta_{ij} \log\left[\frac{\gamma_0 \lambda e^{-\lambda t_i} e^{BZ_j}}{1-e^{-\lambda t_i}}\right] + \log\{\exp(-\gamma_0 e^{BZ_j} \ln(1-e^{-\lambda t_i}))\},$$

$$Log\ L = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left[\delta_{ij} \log(\gamma_0 \lambda e^{-\lambda t_i} e^{BZ_j}) - \log(1-e^{-\lambda t_i})\right] - \gamma_0 e^{BZ_j} \ln(1-e^{-\lambda t_i}),$$

$$Log\ L = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left[\delta_{ij} \log(\gamma_0 \lambda) - \delta_{ij}\lambda t_i + \delta_{ij} BZ_j) - \delta_{ij} \log(1-e^{-\lambda t_i})\right] - \gamma_0 e^{BZ_j} \ln(1-e^{-\lambda t_i}),$$

$$Log\ L = \sum_{j=1}^{J}\left[r_j \log(\gamma_0 \lambda) - \lambda\sum_{i=1}^{n_j} t_{ij} + r_j BZ_j \sum_{i=1}^{n_j}(1-e^{-\lambda t_{ij}}) - \gamma_0 e^{BZ_j}\sum_{i=1}^{n_j} \ln(1-e^{-\lambda t_{ij}})\right], \qquad (4.5)$$

where $r_j = \sum_{i=1}^{n_j} \delta_{ij}$

Fisher information matrix is obtained through evaluation the negative expected values of the corresponding second order partial derivatives of the natural logarithm of the likelihood function with respect to the parameters (λ, B). The maximum likelihood estimators of the model parameters are then asymptotically normally distributed with the variance covariance matrix V given by the inverse of the following Fisher information matrix.

$$V = E\begin{bmatrix} \frac{-\partial^2 \log L}{\partial \lambda^2} & \frac{-\partial^2 \log L}{\partial \lambda \partial B} \\ \frac{-\partial^2 \log L}{\partial B \partial \lambda} & \frac{-\partial^2 \log L}{\partial B^2} \end{bmatrix} \qquad (4.6)$$

$$\frac{\partial LogL}{\partial \lambda} = \sum_{j=1}^{J}\left[r_j \lambda^{-1} - \sum_{i=1}^{n_j} t_i - \sum_{i=1}^{n_j} \frac{t_i e^{-\lambda t_i}}{1-e^{-\lambda t_i}} - \sum_{i=1}^{n_j} \frac{\gamma_0 e^{BZ_j} t_i e^{-\lambda t_i}}{1-e^{-\lambda t_i}}\right]$$

$$\frac{\partial^2 LogL}{\partial \lambda^2} = \left[-r_j \lambda^{-2} + \sum_{i=1}^{n_j} \frac{t^2{}_i e^{-\lambda t_i}}{(1-e^{-\lambda t_i})^2} + \sum_{i=1}^{n_j} \frac{\gamma_0 e^{BZ_j} t^2{}_i e^{-\lambda t_i}}{(1-e^{-\lambda t_i})^2}\right] \qquad (4.7)$$

$$\frac{\partial LogL}{\partial B} = r_j Z_j - \gamma_0 e^{BZ_j} Z_j \sum_{i=1}^{n_j} \ln(1-e^{-\lambda t_i})$$

$$\frac{\partial^2 LogL}{\partial B^2} = \left[-r_j \gamma_0 e^{BZ_j} Z_j^2 \sum_{i=1}^{n_j} \ln(1-e^{-\lambda t_i}) \right] \tag{4.8}$$

$$\frac{\partial LogL}{\partial \lambda \partial B} = -\sum_{i=1}^{n_j} \frac{\gamma_0 Z_j t_i e^{BZ_j} e^{-\lambda t_i}}{(1-e^{-\lambda t_i})}$$

$$\frac{\partial^2 LogL}{\partial \lambda \partial B} = \left[-\sum_{i=1}^{n_j} \frac{\gamma_0 Z_j t_i e^{BZ_j} e^{-\lambda t_i}}{(1-e^{-\lambda t_i})} \right] \tag{4.9}$$

For simplicity of notation, let

$$V^{-1} = \begin{bmatrix} D_\lambda & D_{\lambda B} \\ D_{B\lambda} & D_B \end{bmatrix} \tag{4.10}$$

To derive a formula the variance of $\hat{B}$ (maximum likelihood estimate of B)

$$V^{-1} = \begin{bmatrix} nD_\lambda & nD_{\lambda B} \\ nD_{B\lambda} & nD_B \end{bmatrix}$$

Suppose that $\hat{B}$ is maximum likelihood estimate of B and $\hat{\lambda}$ is maximum likelihood estimate of λ. Then, the variance- covariance matrix of $(\hat{B}, \hat{\lambda})$ is

$$V(\hat{B}) = \frac{nD_\lambda}{\left[nD_\lambda (nD_B) - n(D_{\lambda B})^2\right]} = \frac{D_\lambda}{n\left[D_\lambda D_B - (D_{\lambda B})^2\right]}$$

$$= \frac{1}{n} W, \tag{4.11}$$

where $$w = \frac{D_\lambda}{\left[D_\lambda D_B - (D_{\lambda B})^2\right]} \tag{4.12}$$

5.Tests of the Hypotheses about B:

When the sample size is not predetermined and can be determined to satisfy required levels of α and β risks will be considered. Suppose that it is desired to test the null hypotheses,

$$H_0 : B = B_0 \qquad \& \quad H_1 : B = B_1, \tag{5.1}$$

with level of significance equals α_0 and probability of type II error equals β_0. The required sample size n satisfying these two conditions is developed below.

The variance of the maximum likelihood estimator $\hat{B}$ of B is $\frac{w}{n}$, where n is the cohort size, and w is given by equation (4.12). It follows from large sample theory that

$$Z = \frac{(\hat{B} - B)}{\sqrt{w/n}} \approx N(0,1)$$

Assume that the test decision rule for the testing situation given by equation (5.1)accepts H_0 if $\hat{B} \geq C$ and reject H_0 if $\hat{B} < C$, for some specified number C. then

$$\alpha_0 = P(\text{rejecting } H_0 \mid H_0) = P(\hat{B} < C \mid H_0) = P\left[\frac{\sqrt{n}(\hat{B} - B_0)}{\sqrt{w_0}} < \frac{\sqrt{n}(C - B_0)}{\sqrt{w_0}}\right],$$

where $w_0 = w\ (B = B_0, \lambda, \gamma_0)$ is the value of w given by equation (5.1) evaluated under H_0. Hence

$$Z_{1-\alpha} = \frac{\sqrt{n}(C - B_0)}{\sqrt{w_0}} \tag{5.2}$$

$$\beta_0 = P(\text{accepte } H_0 \mid H_1) = P(\hat{B} \geq C \mid H_1) = P\left[\frac{\sqrt{n}(\hat{B} - B_1)}{\sqrt{w_1}} > \frac{\sqrt{n}(C - B_1)}{\sqrt{w_1}}\right],$$

where $w_1 = w\ (B = B_0, \lambda, \gamma_0)$. Hence

$$Z_{\beta_0} = \frac{\sqrt{n}(C - B_1)}{\sqrt{w_1}} \tag{5.3}$$

From equations (5.2) and (5.3), we obtain

$$n = (B_0 - B_1)^{-2}\left[Z_{\beta_0}\sqrt{W_1} - Z_{1-\alpha_0}\sqrt{W_0}\right]^2$$

Suppose that the lifetime t of individuals under study follows the EE distribution with the following proportional hazard model having three parameters (λ, γ_0, B)

$$r(t; \gamma, \lambda \mid Z_j) = \frac{\gamma_0 \lambda e^{-\lambda t}}{1 - e^{-\lambda t_i}} e^{BZ_j}, \tag{5.4}$$

where λ *and* γ_0 are the parameters of the EE distribution, Z_j is the j the level of a covariate Z which reflects the dose (or concentration), j=1,…,J and B is the parameter of the covariate Z.

Suppose that it is desired to test the null hypothesis

H_0 : B = B_0 versus H_1 : B = B_1 < B_0,

with specified level of significance ($\alpha = \alpha_0$) and the power ($1 - \beta = 1 - \beta_0$).

Suppose that $\gamma = \gamma_0$ and the censoring time for all individuals is T_c . The required sample size to satisfy the requirements on the significance level and power is:

$$n = (B_0 - B_1)^{-2}\left[Z_{\beta_0}\sqrt{W_1} - Z_{1-\alpha_0}\sqrt{W_0}\right]^2 , \tag{5.5}$$

where W_0 is the value of W evaluated under H_0 and C_I is the value of W evaluated under H_1.

In any cohort study while determining its required sample size, say n, the investigator faces the problem of choosing the appropriate length of follow-up period, (or censoring T_c) needed to obtain the prespecified power. Thus, the researcher is advised to use the following algorithm to find the value of n which gives the required power.

Step 1: Determine the null hypothesis (H_0) and the alternative hypothesis (H_1).

Step 2: Choose an appropriate value for $T_c = 0.5(0.5)3$.

Step 3: Given the values of $\lambda = 0.2, 0.4, 1.0., \gamma_0 = 0.1.$, $B_0 = 0, B_1 = -0.04., T_c$ and $Z_j = 1, 2, 3$. Compute the variance- covariance matrix using equation (4.7), (4.8) and (4.9).

Step 4: Compute W using equation (4.12) under the null hypothesis H_0, and it is denoted by W_0. Compute W using equation (4.12) under the alternative hypothesis H_1 and it is denoted by W_1.

Step 5: Determine the value of α to specify the degree of precision and β to determine the required power. Use equation (5.5) to compute the sample size to satisfy the power requirement.

6. Illustrative Example :

Let us consider the example given in Taulbee and Symons (1983). In this example the time t, indicates years until death after the tumor implant. Taking to follow the EE distribution, we suppose that all individuals are given implants and that the individuals are divided into three groups according to whether they will receive doses of 0 , 10 or 20 units of a compound which is believed to slow the spread of the tumor. With units of dose as the covariate, Z_j , the hazard rate is expressed (5.4).

$$r(t;\gamma,\lambda|Z_j) = \frac{\gamma\lambda e^{-\lambda t}}{1-e^{-\lambda t_i}} e^{BZ_j}$$

In all examples given below, we use $\theta_j = \frac{1}{3}$, j = 1, 2, 3, J= 3 (i.e. $n_j = \frac{n}{3}$, j = 1, 2, 3)

Case 1: let $\lambda = 0.2, \gamma = 0.1$ *and* $T_c = 0.5$ year. The hypothesis is H_0: $B = B_0 = 0$ and

H_1: $B = B_1 < B_0$ $\quad B_1 = -0.04$

The censoring time T_c (or follow-up period), is the same for all individuals, and equal group size are used, j = 1, 2, and 3, where $Z_1 = 0$, $Z_2 = 10$ and $Z_3 = 20$. Under H_0 : $B_0 = 0$, we get $W_0 = 0.01$. Under the alternative hypothesis H_1 where $B = B_1 = -0.04$, we have $W_1 = 0.032$. Using (5.5) with $B_0 = 0$, and $B_1 = -0.04$, we obtain the following values of n under different values of α, β

a) If $\alpha = 0.10$ and $(1-, \beta) = 80\%$ the sample size equal 50.

b) If $\alpha = 0.05$ and $(1-, \beta) = 80\%$ the sample size equal 63.

c) If $\alpha = 0.01$ and $(1-, \beta) = 80\%$ the sample size equal 93.

Case 2 : suppose that $\lambda = 0.2, \gamma = 0.1$ *and* $T_c = 1$ year.

Under H_0 : $B_0 = 0$, we get $W_0 = 0.022$. Under the alternative hypothesis H_1 where $B = B_1 = -0.04$, we have $W_1 = 0.063$. Using (5.5) with $B_0 = 0$, and $B_1 = -0.04$, we obtain the following values of n under different values of α, β

d) If $\alpha = 0.10$ and $(1-, \beta) = 80\%$ the sample size equal 102.

e) If $\alpha = 0.05$ and $(1-, \beta) = 80\%$ the sample size equal 131

f) If $\alpha = 0.01$ and $(1-, \beta) = 80\%$ the sample size equal 193.

Table (1): Total sample size for three groups, the study period T_c ranging from 0.5 to 3 years for $\lambda = 0.2, \gamma = 0.1$, power ranging from 80% to 99% and level of significance α equal 0.10, 0.05 and 0.01.

T_c / $1-\beta$ α	T_c=0.5	T_c=1	T_c=1.5	T_c=2	T_c=2.5	T_c=3
.80 .10	50	102	161	293	361	716
.05	63	131	208	306	484	1010
.01	93	193	313	473	756	1661
.90 .10	76	162	245	352	529	1008
.05	93	196	297	447	675	1352
.01	127	272	420	625	992	2093
.95 .10	110	217	330	494	729	1254
.05	129	256	400	599	900	1622
.01	171	342	540	812	1260	2445
.99 .10	225	370	542	784	1089	1482
.05	206	420	625	915	1296	2235
.01	258	529	798	1190	1722	3187

Table (2): Total sample size for three groups, the study period T_c ranging from 0.5 to 3 years for $\lambda=0.4, \gamma=0.1$, power ranging from 80% to 99% and level of significance α equal 0.10, 0.05 and 0.01.

T_c / 1−β	α	T_c=0.5	T_c=1	T_c=1.5	T_c=2	T_c=2.5	T_c=3
.80	.10	41	80	121	156	218	256
	.05	52	100	163	210	280	333
	.01	76	150	240	315	430	518
.90	.10	68	116	203	262	324	400
	.05	81	163	256	324	400	505
	.01	110	225	352	456	575	716
.95	.10	90	176	281	359	452	540
	.05	110	210	342	430	540	650
	.01	144	280	452	583	742	900
.99	.10	156	297	410	546	729	885
	.05	175	342	484	663	841	1024
	.01	217	430	612	841	1089	1332

Table (3): Total sample size for three groups, the study period T_c ranging from 0.5 to 3 years for $\lambda=1, \gamma=0.1$, power ranging from 80% to 99% and level of significance α equal 0.10, 0.05 and 0.01.

T_c / 1−β	α	T_c=0.5	T_c=1	T_c=1.5	T_c=2	T_c=2.5	T_c=3
.80	.10	33	60	77	90	95	101
	.05	42	77	100	116	121	128
	.01	60	110	150	156	182	190
.90	.10	56	95	126	144	160	164
	.05	68	115	156	175	189	197
	.01	90	156	217	240	264	274
.95	.10	81	121	175	203	210	225
	.05	95	144	216	240	248	264
	.01	121	189	280	315	333	351
.99	.10	138	226	297	333	365	372
	.05	157	257	342	380	417	426
	.01	193	315	430	473	523	535

Discussion and Recommendations

The method for sample size determination for follow-up studies of Exponentiated exponential distribution is based on the maximum likelihood estimators of the parameters are asymptotically normally distributed with variance- covariance matrix obtained by inverting the fisher information matrix. The following conclusions are obtained from table (1), (2) and (3).

a) The power (1- β) increases, the sample size (n) increases.

b) If the level of significance (α) decreases the sample size (n) increases.

c) If the follow-up period T_e increases the sample size (n) increases.

d) If the parameter λ increases the sample size (n) decreases.

References

1. Abd-Elfattah,A.M. and Bakoban,R.A.(2003),"A study on the estimation of sample size for generalized gamma distribution" *The Scientific Journal for Economic and Commerce, Faculty of Commerce, Ain Shames University,* ***Vol.2, 17-23.***

2. Ashour, S. K., Abd-Elfattah, A.M.and Mahmoud, M.R.(1996),"Estimation of sample size for the Burr failure model." *The Annual conference in Statistics, Computer Science and Operations Research, ISSR,Cairo Uuiv.,****31(2),18-26.***

3. Ashour, S. K. and Shalaby, O.A.(1983),"Estimating sample size with Weibull failure." *Math. Operationsforsch.U.Statist. Ser. Statist., Berlin, 2,263-268.*

4. Chiange C. (1961), "The follow-up study with the consideration of competing risks". ***Biometrics, 29, 57-78.***

5. Cox, D. (1972), "Regression models and life-tables" *J.R. statistic. Soc.* ***B, 34, 187-220.***

6. Cox, D. and Oakes, D. (1984), "*Analysis of survival data*" *New York: Chapman and Hall.*

7. Feigl, P. and Zelen, M.(1965), "Estimation of Exponential survival probabilities with concomitant information " *Biometrics 21, 826- 838.*

8. Efron and Johnstone (1990), "Fisher's Information in Terms of the Hazard Rate"*Ann. Statist.* ***18, Number 1, 38-62.***

9 George, S. and Desu, M. (1974),"Planning the size and duration of a clinical trial studying the time to some critical event" *Journal of chronic diseases 27, 15-24.*

10. Gokhale, D.V.(1972),"A formula for sample size Required for Specified Tolerance." *The American Statistician, 26, 21.*

11. Guenther, W.C.(1970)," Determination of sample size for distribution free Tolerance limits." *The American Statistican, 24, 44-45.*

12. Gupta, R.D., and Kundu, D. (1999)," generalized exponential distributios"Australian and New Zealand *Journal of Statistics, 41(2)173-188*

13. Gupta, R.D., and Kundu, D. (2001a)," Exponentiated exponential distribution:statistical infernces, an alternative to gamma and Weibul distributions" *Biometric Journal 43,117-130.*

14. Gupta, R.D., and Kundu, D. (2001b)," Generalized exponential distribution: different methods of estimation." *Journal of Statistical computation and Simulation, 69,315-337.*

15. Gupta, R.D., and Kundu, D. (2002)," Generalized exponential distributions: Statistical inferences." *Journal of Statistical theory and Applications 1,101-118.*

16. Gupta, R.D., and Kundu, D. (2003)," Closeness of gamma and generalized exponential distribution." *Communication in Statistics-theory and Methods 32,705-721.*

17. Gupta, R. D. and Kundu, D. (2006), " On the comparison of Fisher information of the Weibull and GE distributions", *Journal of Statistical Planning and Inference, 136, 3130–3144.*

18. Harman, A.J. (1967)," Note on wilks tolerance limit sample size." *Sankhya, A, 30,215-219.*

19. Johnson, N.L. (1962),"Estimation of sample." *Tehnometrics, 4, 59-67.*

20. Kung, J.L.(1993)," Sample size determination for cohort studies under an exponential covariate model with group data " *Biometrics 49,733-778.*

21. LawLess, J. E. (1982), "*Statistical Models and Methods for Lifetime data*". **John Wiley Sons.**

22. Nadarajah, S.and Kotz, S.(2006)," The exponentiated type distributions ". *Acta Appl. Math.*, 92, 97-111.

23. Nanda, A.K. and Gupta,R.D.(2001), "Some properties of reversed hazard rate function" *Statistical Methods, 3,108-124.*

24. Saleh, S. M. (2002)," The Relationship between sample size and the length of the Follow –up Period" *Ph.D. thesis, Al Azhar University.*

25. Taulbee, J. D. and symons, M. J. (1983), "Sample size and duration for cohort studies of survival time with covariables". *Biometrics 39,351-360*