Modified Box-Cox Transformation and Manly Transformation with Failure Time Data

Lakhana Watthanacheewakul

Backgrounds

- Many procedures require data to be approximately normal.
- A transformation that transforms the data set to achieve normality is used.

Data Transformation

- **Based on the relationship between the standard deviation and the mean**

Relationship between σ and μ	Transformation
$\sigma \propto \mu^{1/2}$	Square root
$\sigma \propto \mu$	Logarithmic
$\sigma \propto \mu^{3/2}$	Reciprocal square root
$\sigma \propto \mu^{2}$	Reciprocal

Source: Montgomery, 2001: 84.

Transformations for Specific Distributions

For Example, the square root transformation is used for Poisson data, the logarithmic transformation for lognormal data and the arcsine transformation for binomial data expressed as fractions.

- **A family of transformations**
  Box and Cox (1964)

$$Y = \begin{cases} \dfrac{X^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln X, & \lambda = 0 \end{cases}$$

$$Y = \begin{cases} \dfrac{[X+c]^{\lambda} - 1}{\lambda} & ,\lambda \neq 0 \\ \ln[X+c] & ,\lambda = 0 \end{cases}$$

- where c is translation constant

Cautions for the Box-Cox Transformation

- John and Draper (1980) showed that the Box-Cox Transformation was not satisfactory even when the best value of transformation parameter had been chosen.

- Doksum and Wong (1983) indicated that the Box-Cox transformation should be used with caution in some cases such as failure time and survival data.

- **Schlesselman (1971)**

$$Y = \begin{cases} \dfrac{X^{\lambda} - c^{\lambda}}{\lambda}, & \lambda \neq 0 \\ \ln(X / c), & \lambda = 0 \end{cases}$$

- where c is an arbitrary positive constant in the measurement units of variable X.

- **Manly(1976)**

$$Y = \begin{cases} \dfrac{\exp(\lambda X) - 1}{\lambda}, & \lambda \neq 0 \\ X, & \lambda = 0 \end{cases}$$

The Modified Box and Cox transformation

- Yeo and Johnson (2000)

$$Y = \begin{cases} \dfrac{(X+1)^{\lambda} - 1}{\lambda} & , x \geq 0,\ \lambda \neq 0 \\ \ln(X+1) & , x \geq 0,\ \lambda = 0 \\ -\dfrac{(-X+1)^{2-\lambda} - 1}{2-\lambda} & , x < 0,\ \lambda \neq 0 \\ -\ln(-X+1) & , x < 0,\ \lambda = 0 \end{cases}$$

For Example

- Rahman M. and Pearson, L.M. (2007). A Note on the Modified Box-Cox Transformation. **Festschrift in honor of Distinguished Professor Mir Masoom Ali on the occasion of his retirement** , May 18-19. 106-115.
- Abbasi, B., Niaki, S.T.A. and Seyedan, S.E. (2011). A Simple Transformation Method in Skewness Reduction. **IJE Transactions A:Basics**. 24(2): 169-175.

Failure Time Data

- Gamma Distribution

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & , x \geq 0,\ \alpha,\beta > 0 \\ 0 & , x < 0 \end{cases}$$

Failure Time Data

- **Exponential distribution**

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\left(\frac{x}{\beta}\right)} & , x \geq 0; \beta > 0 \\ 0 & , x < 0 \end{cases}$$

Comparisons of Several Population Means

- The probability density function of each transformed observation is in the form

$$f(y_{ij} \mid \mu_i, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{ -\frac{1}{2\sigma^2}(y_{ij} - \mu_i)^2 \right\}$$

Estimation of Transformation Parameter for Modified Box and Cox Transformation

- The likelihood function in relation to the original observations is given by

$$L(\mu_i, \sigma^2, \lambda | x_{ij}) =$$

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[\frac{(x_{ij}+1)^{\lambda} - 1}{\lambda} - \mu_i \right]^2 \right\} . J(y;x) \quad \text{where}$$

$$J(y;x) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \left| \frac{\partial y_{ij}}{\partial x_{ij}} \right| .$$

Transformation Parameter in Modified Box -Cox transformation

- The maximum likelihood estimate of transformation parameter is obtained by solving the likelihood equation

$$\frac{d}{d\lambda}\ln L(\lambda) = \frac{-n\left[\begin{array}{l}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}+1)^{2\lambda}\ln(x_{ij}+1)- \\ \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i}(x_{ij}+1)^{\lambda}\right)\left(\sum_{j=1}^{n_i}x_{ij}^{\lambda}\ln(x_{ij}+1)\right)\end{array}\right]}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}+1)^{2\lambda}-\sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i}(x_{ij}+1)^{\lambda}\right)^{2}} + \frac{n}{\lambda}+\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln(x_{ij}+1)=0.$$

Estimation of Transformation Parameter for Manly Transformation

- The likelihood function in relation to the original observations is given by

$$L(\mu_i, \sigma^2, \lambda | x_{ij}) =$$

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[\frac{\exp(\lambda x_{ij}) - 1}{\lambda} - \mu_i \right]^2 \right\} . J(y;x) \quad \text{where}$$

$$J(y;x) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \left| \frac{\partial y_{ij}}{\partial x_{ij}} \right| .$$

Transformation parameter in Manly Transformation

- The maximum likelihood estimate of transformation parameter is obtained by solving the likelihood equation

$$\frac{d}{d\lambda}\ln L(\lambda) =$$

$$\frac{-n\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} e^{2\lambda x_{ij}} x_{ij} - \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} e^{\lambda x_{ij}}\right)\left(\sum_{j=1}^{n_i} e^{\lambda x_{ij}} x_{ij}\right)\right]}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} e^{2\lambda x_{ij}} - \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} e^{\lambda x_{ij}}\right)^2}$$

$$+\frac{n}{\lambda} + \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} = 0.$$

Check Validity of Assumption

- The Kolmogorov-Smirnov Test
- The Levene Test

Simulation Study for the Gamma Data

- The possible value for study is set as follows:

k = number of the Gamma populations = 3,

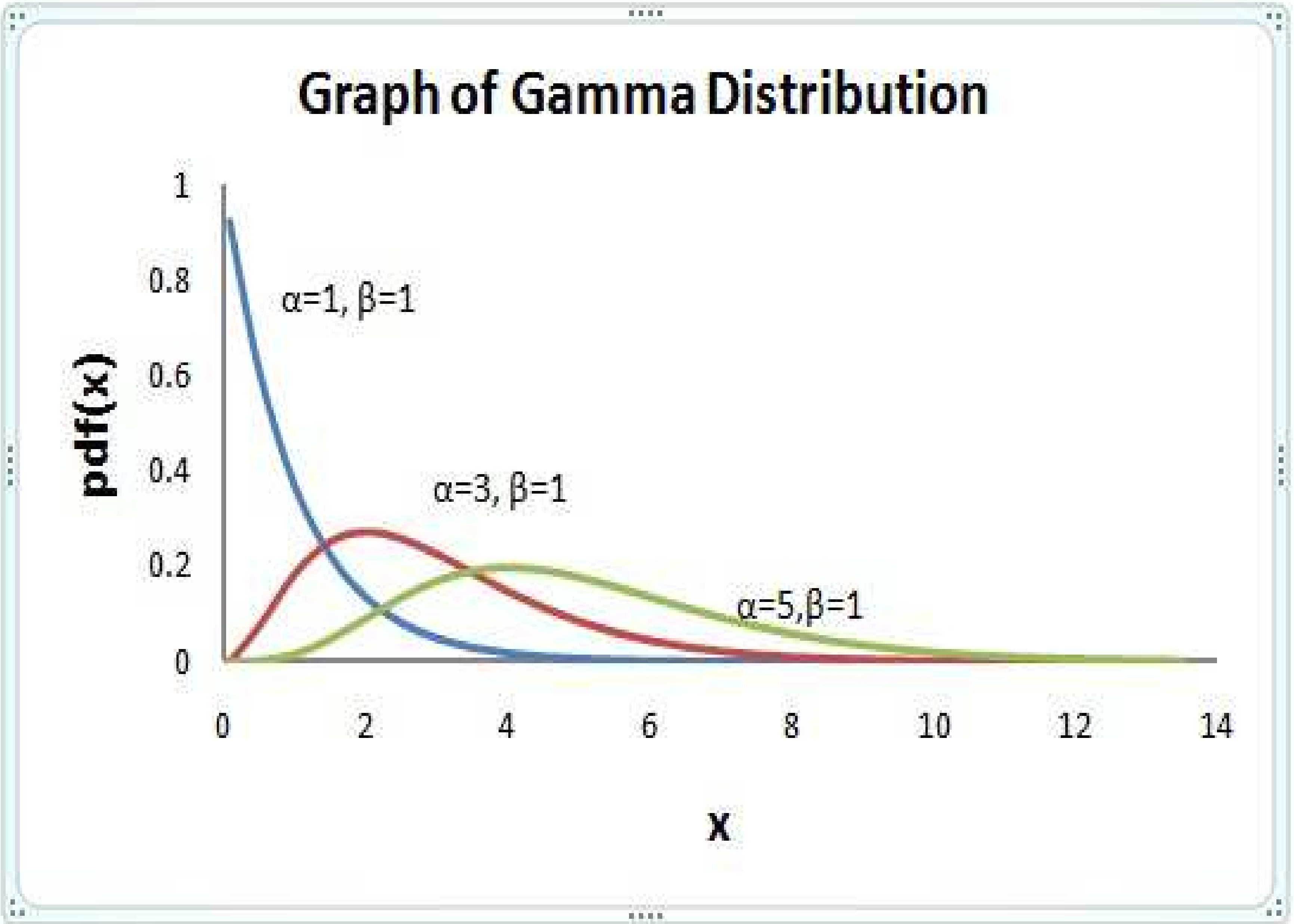
n_i = sample size from the i th Gamma population is between 5 and 90,

β_i= scale parameter of the i th Gamma population is between 1 and 3 ,

α_i = shape parameter of the i th Gamma population is between 1 and 5

The Results

- The results of the goodness- of-fit tests and the tests of homogeneity of variances with 1,000 replicated samples of various sizes are as follows



Graph of Gamma Distribution
pdf(x)
1
0.8
0.6
0.4
0.2
0
α=1, β=1
α=3, β=1
α=5,β=1
0
2
4
6
8
10
12
14
x

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=10	Manly	0.7861	0.7858	0.7933
	Modified	0.7866	0.7883	0.7910
n_i=30	Manly	0.6245	0.6563	06958
	Modified	0.6262	0.6640	0.6930

Averages of the p-Values for K-S Test

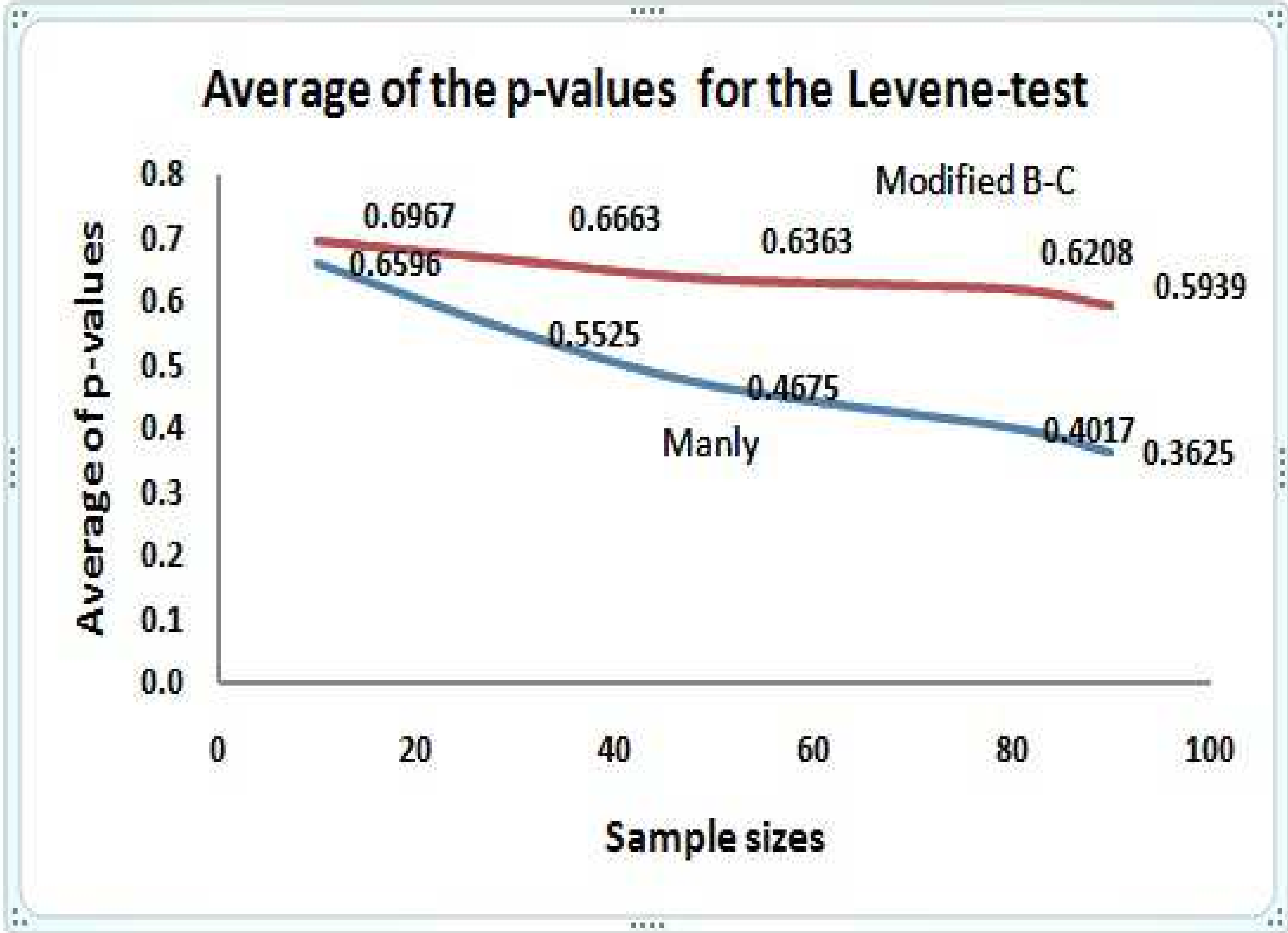
Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=50	Manly	0.5045	0.5427	0.5975
	Modified	0.5077	0.5558	0.5930
n_i=80	Manly	0.3625	0.4124	0.4799
	Modified	0.3656	0.4294	0.4904

Averages of the p-Values for K-S Test

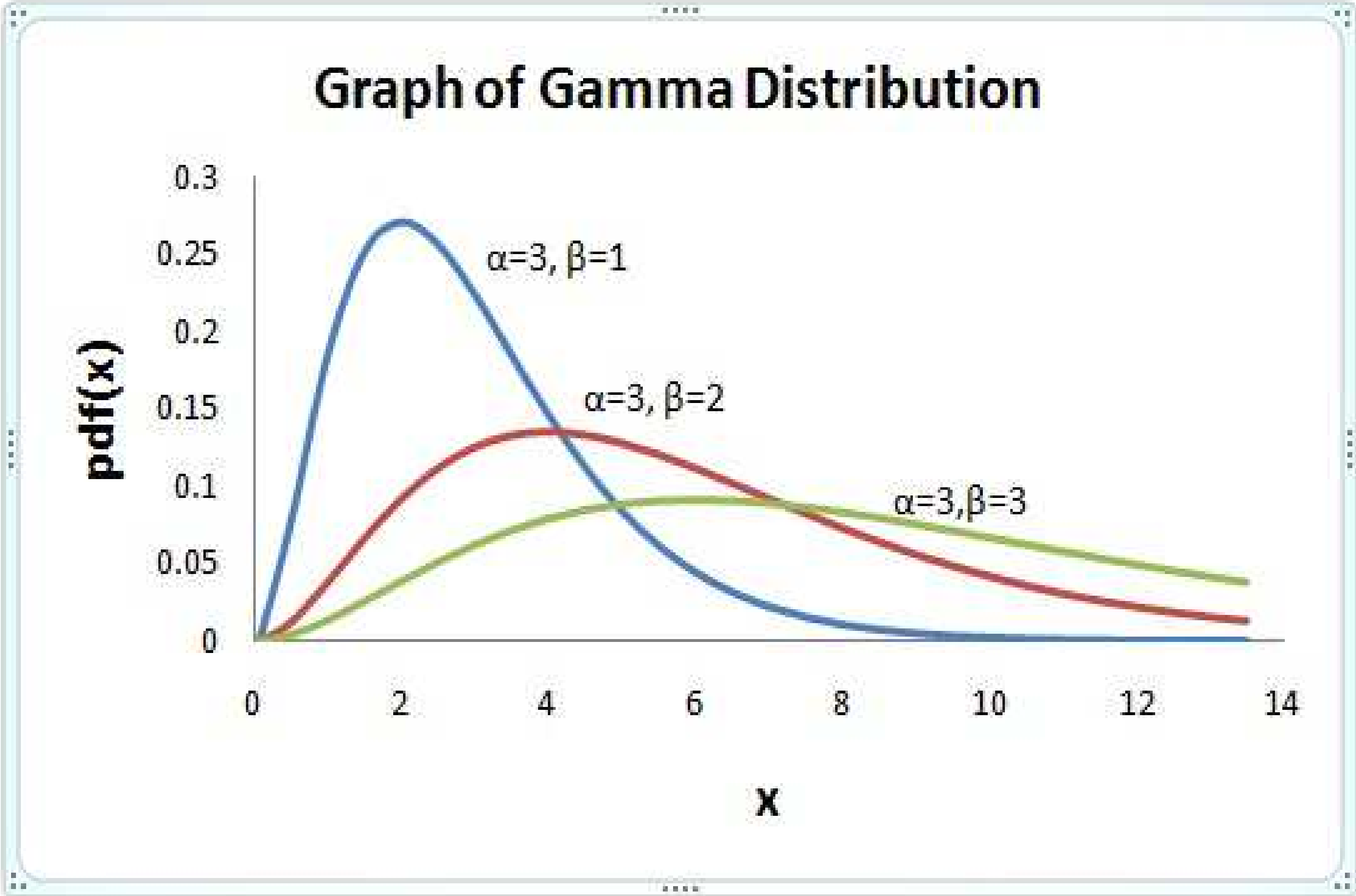
Sample sizes	Transformations	Averages of the p-Values for K-S Test		
$n_i=90$	Manly	0.3398	0.3732	0.4562
	Modified	0.3430	0.3921	0.4509
$n_1=5$, $n_2=10$,	Manly	0.8369	0.7793	0.7589
$n_3=15$	Modified	0.8383	0.7803	0.7566

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_1=5, n_2=15,	Manly	0.8430	0.7495	0.6106
n_3=25	Modified	0.8445	0.7502	0.6114
n_1=10, n_2=30	Manly	0.7809	0.6447	0.5748
n_3=50	Modified	0.7873	0.6476	0.5731



Average of the p-values for the Levene-test
Modified B-C
Manly
0.6967
0.6663
0.6363
0.6208
0.5939
0.6596
0.5525
0.4675
0.4017
0.3625
Average of p-values
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
Sample sizes


Graph of Gamma Distribution
0.3
0.25
0.2
0.15
0.1
0.05
0
pdf(x)
α=3, β=1
α=3, β=2
α=3,β=3
0
2
4
6
8
10
12
14
x

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=10	Manly	0.7758	0.7764	0.7693
	Modified	0.7803	0.7977	0.7707
n_i=30	Manly	0.6358	0.6288	0.6030
	Modified	0.6529	0.6315	0.6085

Averages of the p-Values for K-S Test

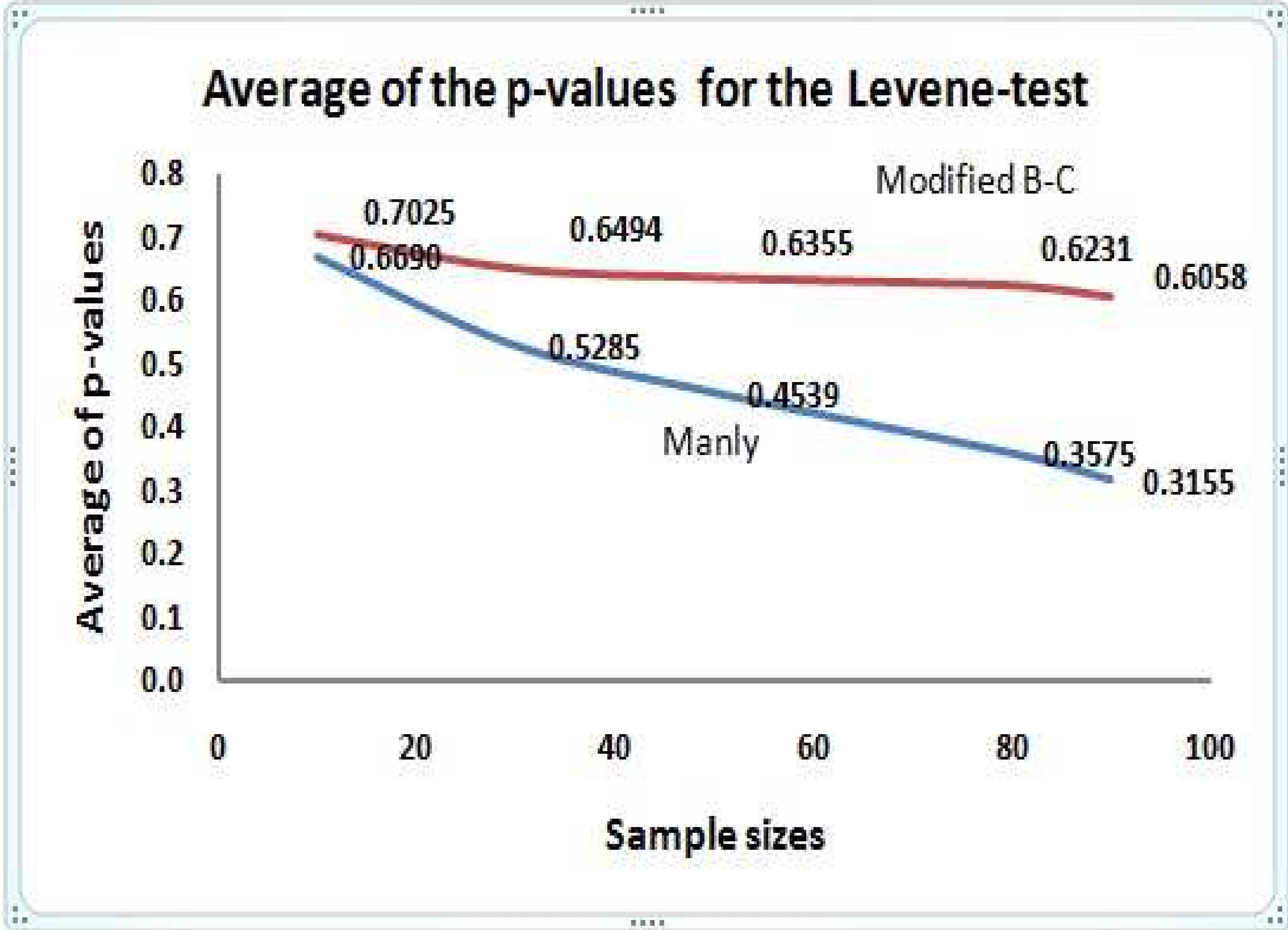
Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=50	Manly	0.5022	0.4897	0.4623
	Modified	0.5251	0.4929	0.4739
n_i=80	Manly	0.3769	0.3499	0.3208
	Modified	0.4057	0.3524	0.3325

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
$n_i=90$	Manly	0.3389	0.3077	0.3961
	Modified	0.3685	0.3105	0.3104
$n_1=5,\ n_2=10,$	Manly	0.8329	0.7940	0.7447
$n_3=15$	Modified	0.8348	0.7941	0.7455

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_1=5, n_2=15,	Manly	0.8407	0.7625	0.6740
n_3=25	Modified	0.8440	0.7624	0.6782
n_1=10, n_2=30	Manly	0.7978	0.6767	0.5257
n_3=50	Modified	0.8054	0.6769	0.5279



Average of the p-values for the Levene-test
Modified B-C
Manly
0.7025
0.6494
0.6355
0.6231
0.6058
0.6690
0.5285
0.4539
0.3575
0.3155
Average of p-values
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
20
40
60
80
100
Sample sizes

Simulation Study for the Exponential Data

- The possible value for study is set as follows:

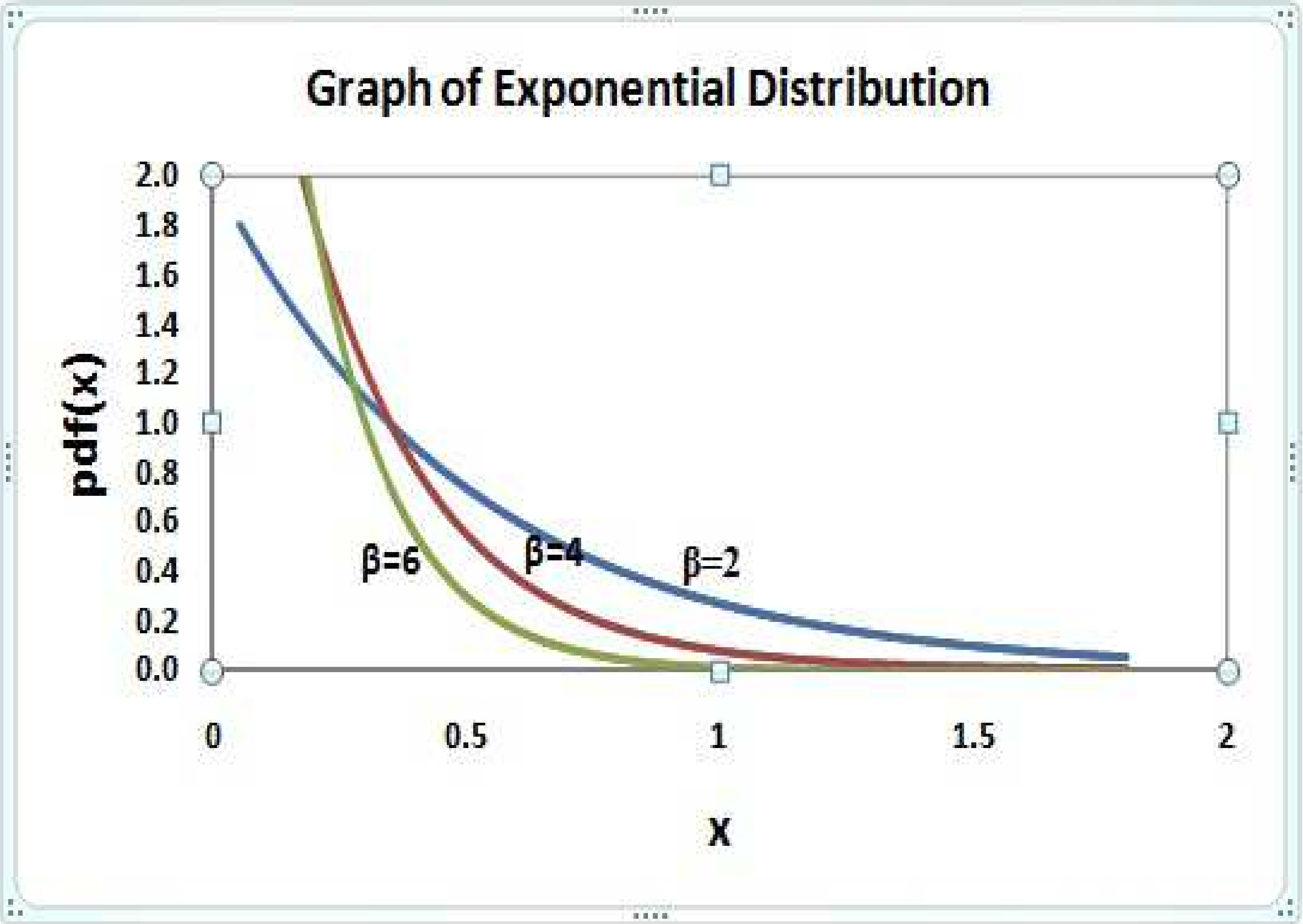
k = number of the Exponential populations = 3,

n_i = sample size from the i th Exponential population is between 5 and 90,

β_i= scale parameter of the i th Exponential population is 2 and 9.

The Results

- The results of the goodness- of-fit tests and the tests of homogeneity of variances with 1,000 replicated samples of various sizes are as follows



Graph of Exponential Distribution
pdf(x)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
β=6
β=4
β=2
0
0.5
1
1.5
2
X

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=10	Manly	0.7083	0.8104	0.8237
	Modified	0.8206	0.8409	0.8381
n_i=30	Manly	0.4432	0.7083	0.6802
	Modified	0.7229	0.7586	0.6858

Averages of the p-Values for K-S Test

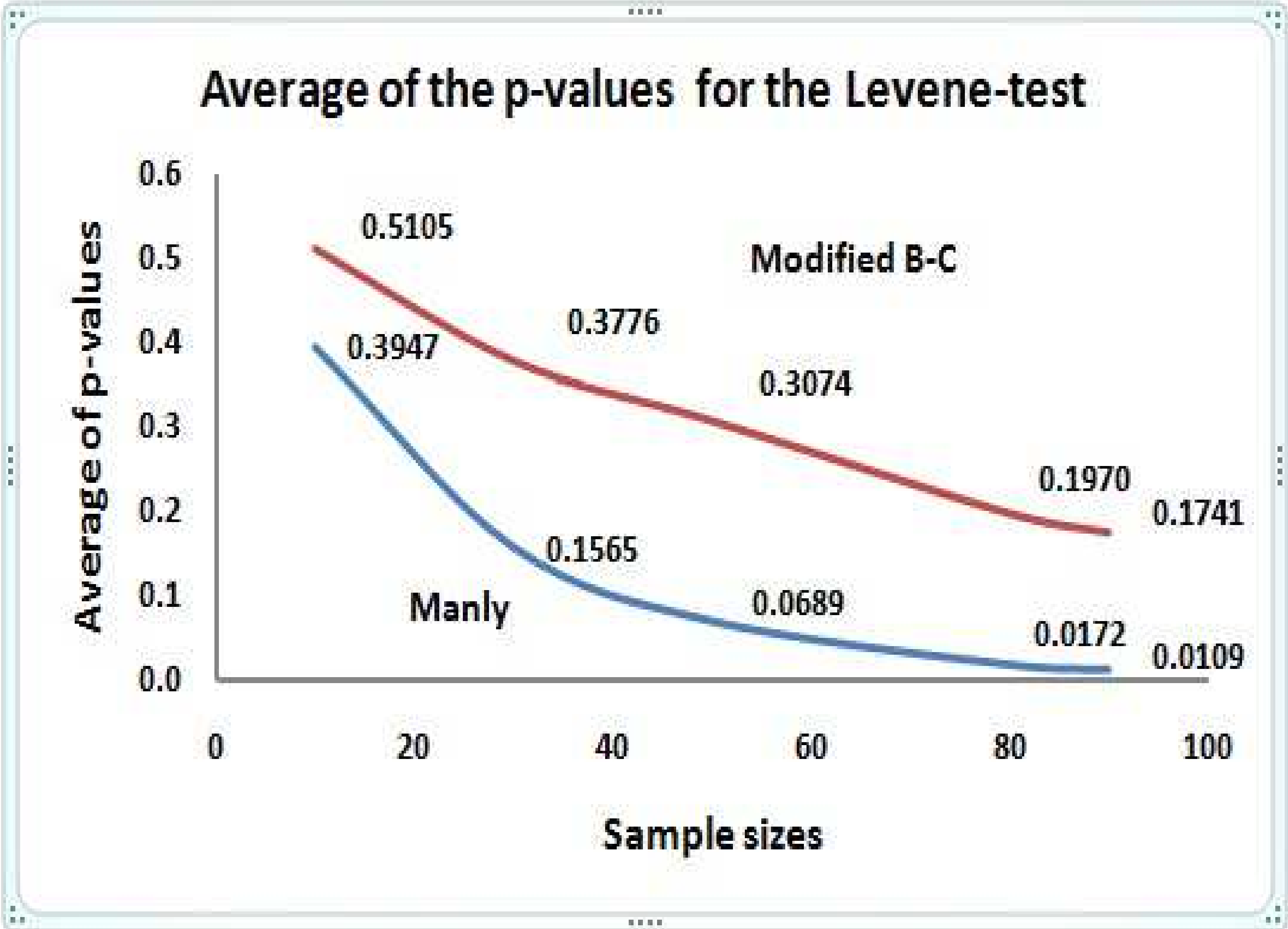
Sample sizes	Transformations	Averages of the p-Values for K-S Test		
$n_i=50$	Manly	0.2987	0.6234	0.5810
	Modified	0.6701	0.7056	0.5921
$n_i=80$	Manly	0.1496	0.4843	0.4443
	Modified	0.5457	0.6254	0.4558

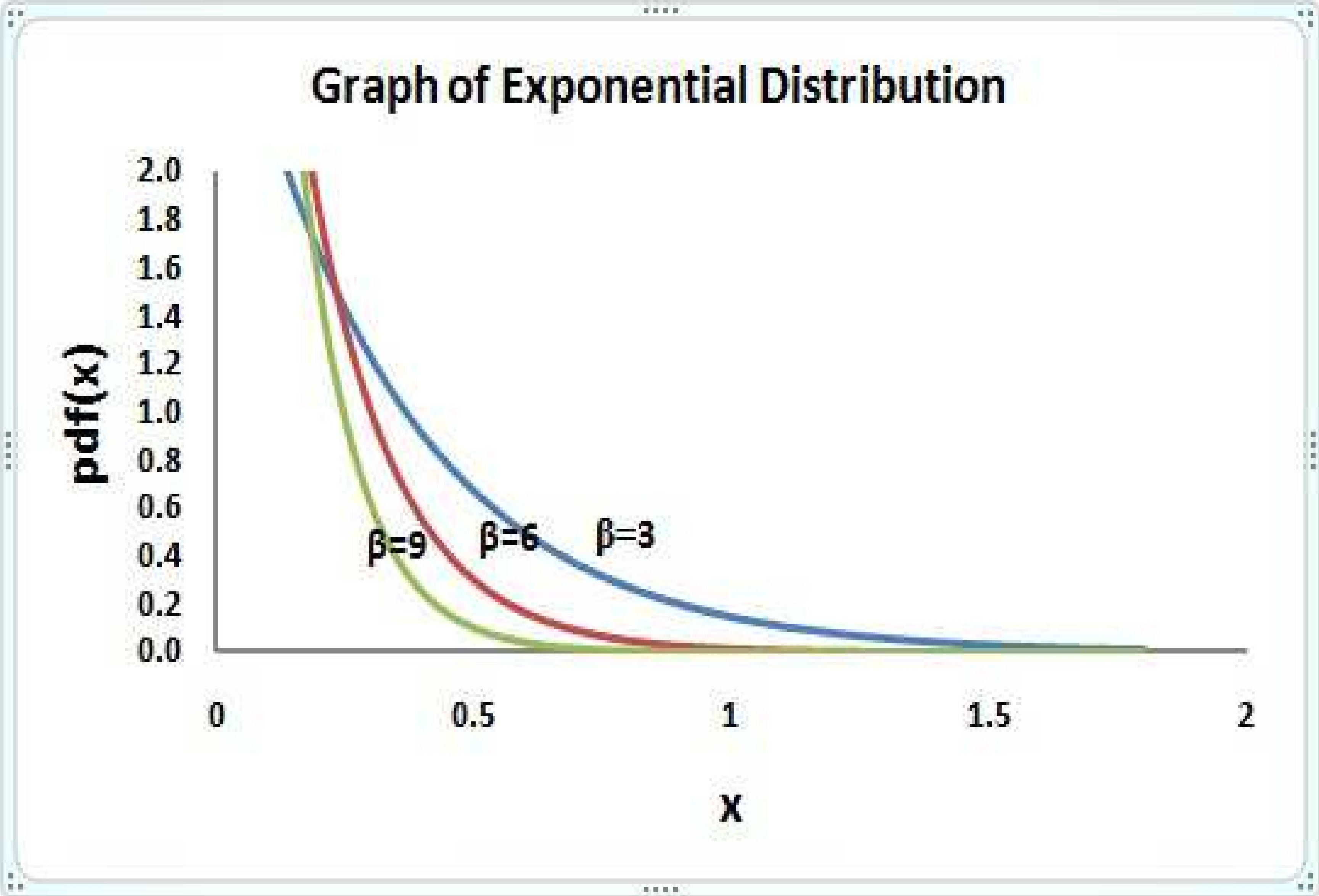
Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=90	Manly	0.1210	0.4596	0.4160
	Modified	0.5246	0.5970	0.4091
n_1=5, n_2=10,	Manly	0.7989	0.7804	0.7669
n_3=15	Modified	0.8414	0.8223	0.8028

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_1=5, n_2=15,	**Manly**	0.8037	0.7252	0.7136
n_3=25	**Modified**	0.8449	0.8111	0.7880
n_1=10, n_2=30	**Manly**	0.6872	0.5985	0.5346
n_3=50	**Modified**	0.7896	0.7640	0.6814


Average of the p-values for the Levene-test
Average of p-values
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
Sample sizes
Modified B-C
Manly
0.5105
0.3776
0.3074
0.1970
0.1741
0.3947
0.1565
0.0689
0.0172
0.0109


Graph of Exponential Distribution
pdf(x)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
β=9
β=6
β=3
0
0.5
1
1.5
2
X

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=10	Manly	0.7042	0.7907	0.8086
	Modified	0.8049	0.8265	0.8258
n_i=30	Manly	0.4407	0.6572	0.6729
	Modified	0.6948	0.7466	0.7239

Averages of the p-Values for K-S Test

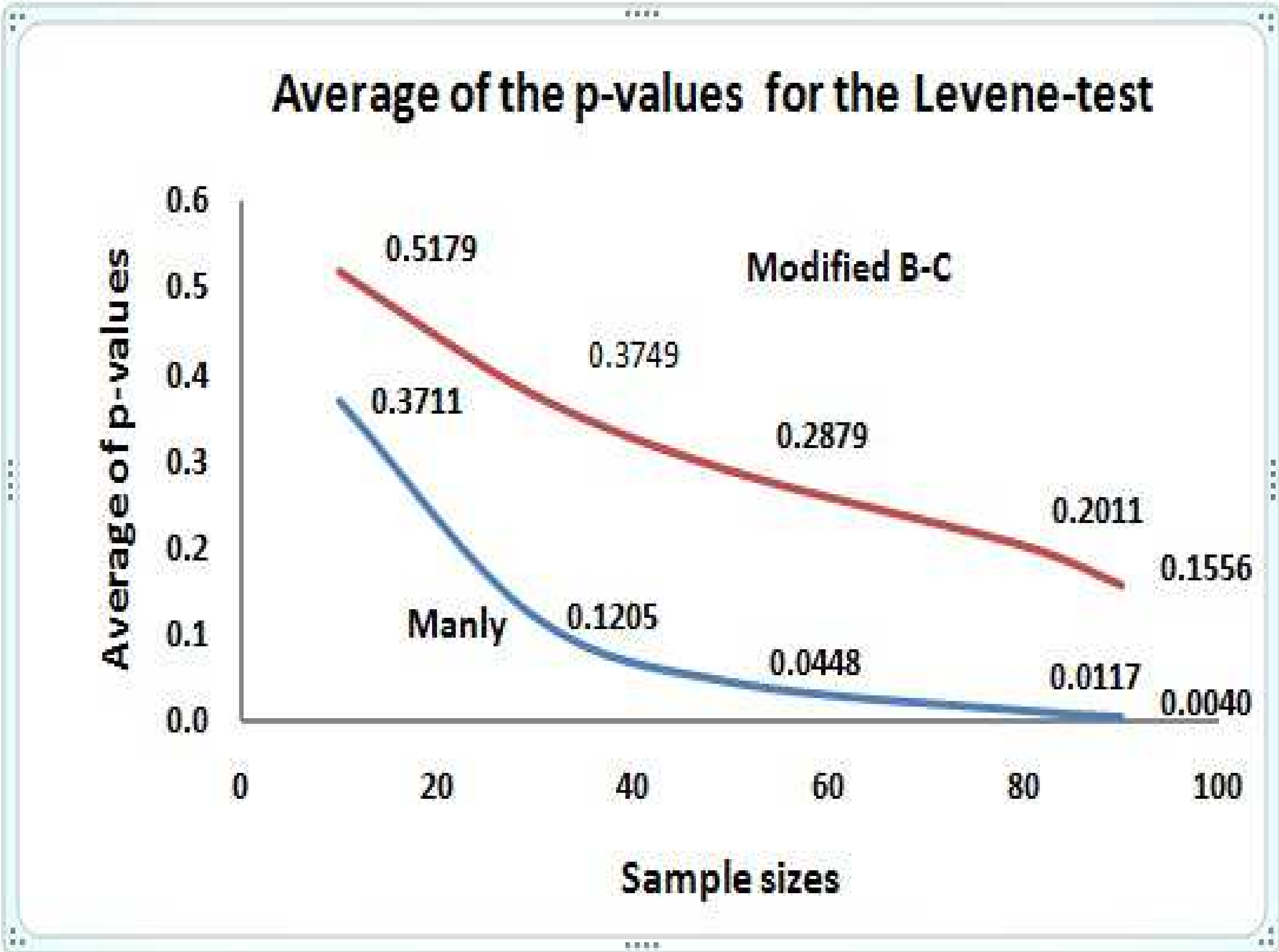
Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=50	Manly	0.2975	0.5552	0.5377
	Modified	0.6154	0.6846	0.6189
n_i=80	Manly	0.1553	0.4347	0.4130
	Modified	0.4945	0.6139	0.5249

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
n_i=90	Manly	0.1254	0.3935	0.3544
	Modified	0.4706	0.5796	0.4740
n_1=5, n_2=10,	Manly	0.8110	0.7851	0.8002
n_3=15	Modified	0.8486	08213	0.8287

Averages of the p-Values for K-S Test

Sample sizes	Transformations	Averages of the p-Values for K-S Test		
$n_1=5$, $n_2=15$,	Manly	0.8070	0.7475	0.7508
$n_3=25$	Modified	0.8434	0.8125	0.8033
$n_1=10$, $n_2=30$	Manly	0.7244	0.6356	0.6413
$n_3=50$	Modified	0.8208	0.7756	07265


Average of the p-values for the Levene-test
Average of p-values
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
Sample sizes
Modified B-C
Manly
0.5179
0.3749
0.2879
0.2011
0.1556
0.3711
0.1205
0.0448
0.0117
0.0040

Conclusions

- Both two transformations can transform the failure time data to correspond with the basic assumptions.
- However, It seems that sample sizes affect on Levene test.

References

- Box, G.E.P. and Cox, D. R. 1964. An Analysis of Transformations (with discussion), **Journal of the Royal Statistical Society, Ser.B**. 26: 211-252.
- Doksum, K. A. and Wong, C. 1983. Statistical Tests Based on Transformed Data. **Journal of the American Statistical Association**. 78: 411-417.
- Gurka, M.J., Edwards, L.J. and Nylander-French, L. Testing Transformations for the Linear Mixed Model. **Computational Statistics & Data Analysis**. 51: 4297-4307.
- John, J. A. and Draper, N. R. 1980. An Alternative Family of Transformations. **Applied Statistics**. 29(2): 190-197.
- Manly, B. F. 1976. Exponential Data Transformations. **Statistician.** 25(1): 37-42.
- Rahman M. and Pearson, L.M. 2007. A Note on the Modified Box-Cox Transformation. **Festschrift in honor of Distinguished Professor Mir Masoom Ali on the occasion of his retirement** , May 18-19. 106-115.

- Schlesselman, J. 1971. Power Families: A Note on the Box and Cox Transformation. **Journal of the Royal Statistical Society, Ser. B**. 33:307-311.
- Tan ,W.D., Gan, F.F. and Chang, T.C. 2004. Using Normal Quantile Plot to Select an Appropriate Transformation to achieve Normality. **Computational Statistics & Data Analysis**. 45: 600-619.
- Yeo, I.K. and Johnson, R. 2000. A New Family of Power Transformations to Improve Normality or Symmetry. Biometrika. 87: 954-959.