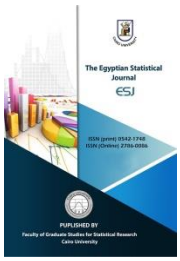
Cairo University

The Egyptian Statistical Journal

Volume (67)-No. 2 -2023
Homepage:.mskas.journals.ekb.eg
Print ISSN 0542-1748– Online ISSN 2786-0086



Estimation of Gamma Distribution Parameters with Incomplete Data

Naglaa Abdelmomeim Morad

Department of Applied Statistics and Econometrics,
Faculty of Graduate Studies for Statistical Research
Cairo University, Egypt

Submitted: 28-08-2023 Revised: 29-11-2023 Accepted: 24-12-2023

ABSTRACT

The missing data problem has been broadly studied in the last few decades. Some researchers studied estimation and hypothesis testing for different distributions .The contributions of the current study was to obtain the estimators parameters of the gamma distribution with missing data for one and two populations. The estimators are obtained using the maximum likelihood method. To compare the suggested maximum likelihood estimators with estimates that might come from various estimation techniques, including the listwise method and the mean imputation method, a simulation study has been conducted. A simulation study with three distinct percentages of missing values in the data sets: 10%, 20%, and 30% as well as three different sample sizes (10, 30, and 50). The estimators' criteria's mean square error (MSE) and the relative absolute biases (RAB) were utilized for comparison. The results demonstrated that the use of mean imputation or the maximum likelihood method for the scale parameter (θ) but the listwise method for the shape parameter (k) is preferred as the percentage of missing data increases. Use the mean imputation approach for the scale parameter (θ) as the data set size grows.

Keywords:

Gamma Distribution, Missing Data, Maximum Likelihood Estimators, Listwise Method, Mean Imputation Method, and Mean Square Error.

1. Introduction

One of the most common problems with data quality is missing values. "Missing data" refers to both the total amount of missing data for a single participant and the total amount of missing data for a specific variable within a data set. Data loss can occur for a variety of reasons, including

application or communication problems, user error in not capturing data, user intention to leave fields empty, and errors in data integration. Statistics has always struggled with how to handle missing data, but recently it has drawn greater attention. The fundamental cause of the current interest in missing data is the challenges presented by surveys and censuses. A remedy, nevertheless, hasn't been agreed upon. As a result, researchers have developed a wide range of techniques to estimate the unknown parameters of various models when there are missing data, including the maximum likelihood method, mean imputation, listwise algorithms, etc. Practically speaking, the processes are presumptions that control how well various missing data strategies work. In many statistical analyses, especially those related to the social sciences, missing data are a persistent and common problem. Therefore, handling missing data is a feature of various statistical software (Acock, 2005). Problems with missing data occur in almost all developmental research projects. A common approach to the issue of missing data is to delete the cases that include those data points. Because it can significantly lower sample size, statisticians have shown that this type of approach to addressing missing data is insufficient. Additionally, by generating results that are not representative of the population, this reduces power and creates bias. Depending on the quantity, causes, patterns, and missingness of the data. Several authors estimated parameters and tested distribution parameters when the data is missing e.g [Poisson, Bernolli, Binomial, Negative binomial, Exponential and Normal distribution…]. Zhao et al, (2009) estimated parameters and tested hypothesis of means of two exponential populations under type I censoring sample when data are missing. Zhao(2012) get the parameter estimation and hypothesis testing on the equality of two negative binomial distribution populations with missing data. Luo (2013) estimated the parameters and tested the hypothesis for two pareto distribution populations with partially missing data. See [Kumar et al. (2017) ,Gupta and Grover (2017) ,Golden et al.(2019) , Nguyen et al.(2021), and Farzandi, et al.(2022)].

The structure of this essay is as follows: The introduction is in Section 1, and an overview of missing data is in Section 2. Gamma Distribution is described in Section 3. Section 4 introduces Estimation of Parameters. Finally, Results and Discussion, and the conclusion are included in sections five and six, respectively.

2. An Overview of Missing Data

This section's objective is to give a general overview of missing data and approaches for handling with it.

2.1 Types of Missing Data

Missing data are errors because the data don't represent the true values of what the set out to measure. Missing values can be of three different types. These types describe relationships between measured variables and the probability of missing data.

2.1.1 Missing not at random (MNAR): Missing data systematically differs from the observed values.

2.1.2 Missing at Random (MAR): Missing data are not randomly distributed, but they are accounted for by other observed variables.

2.1.3 Missing Completely at Random (MCAR): Missing data are randomly distributed across the variable and unrelated to other variables.

See [Baraldi and Enders (2010)].

2.2 Methods of Handling Missing Data

When the data was absent, several authors calculated and evaluated the distribution parameters. Donders et al, (2006) used imputation techniques to handle missing data. Dong and Peng (2013) demonstrated three principled missing data methods: multiple imputation, full information maximum likelihood, and expectation-maximization algorithm.

2.2.1 Listwise Deletion or Complete Case Analysis

The simplest method for handling missing data is listwise deletion. When using this technique, cases are removed from the sample if any of the variables in the analysis that will be done have missing data. This results in a working sample with no missing data, allowing any statistical approach to be used thereafter. Listwise deletion offers two significant statistical qualities in addition to being straightforward and generic. First off, listwise deletion won't skew the parameter estimates if the data are absent entirely at random. The subsample with complete data, if the data are MCAR, is essentially a simple random sample from the main sample. Simple random sampling does not introduce bias, as is well known. Second, the listwise deletion standard error estimates should roughly represent unbiased estimates of the real standard errors. This is crucial because the majority of other conventional approaches' standard error estimations have problems for one reason or another. Listwise deletion may result in biassed parameter estimates if data are absent at random but not entirely at random. See [Pepinsky (2018)].

2.2.2 Mean Imputation

Mean imputation, sometimes referred to as regression-based imputation, is a technique for imputed missing data that uses multiple regression modelling to predict values. A multiple regression model is calculated on the available cases in which the variable with missing data is regressed on other measured variables using all cases with complete data for the regression, for instance, if there are several measured variables within a data set and only one has missing data. The multiple regression equation's predicted scores are then used to fill in any missing data. Since this method uses imputation values from a data analysis model with complete data, it appears to be a great way to handle missing data theoretically.However, there are a number of issues with this approach. First, because all projected values used for imputation fall exactly on the regression line, the imputed data lacks variability. As a result, this strategy generates biassed estimates of variance and covariance, as demonstrated by simulation experiments. Second, a univariate pattern of missing data is required for regression imputation to be straightforward to apply. Regression imputation can be challenging if a data set contains a monotone or random pattern of missing data since multiple regression equations must be created for each distinct pattern. See [Jäger et al. (2021)].

2.2.3 The Maximum Likelihood (ML) Method

Maximum likelihood determines the parameter values that have the maximum likelihood of creating the sample data using all of the available data, both complete and incomplete. In essence, the estimation process repeatedly performs log-likelihood calculations while substituting new population parameter values into the log-likelihood equation at each iteration. The objective of estimating is to determine the specific constellation of estimates that produces the highest log-likelihood and, thus, the best fit to the data. Each distinct combination of parameter estimations produces a different log-likelihood number. The favorable qualities of ML estimators are numerous. They are known to be consistent, asymptotically effective, and asymptotically normal under a variety of circumstances. Consistency means that in large samples, the estimates are roughly unbiased. In order to be efficient, an estimator's true standard errors must be at least as minimal as those of any other consistent estimator. Finally, asymptotic normality denotes that estimates in repeated sampling have a distribution that is roughly normal (again, the approximation gets closer with increasing sample size). Small sample sizes may result in biassed parameter estimates when using ML.

The likelihood function with missing data has the form:

$$L(\theta) = \prod_{i=1}^{n} \big(f(x_i;\theta)\big)^{\delta_i} \quad (1)$$

$$\delta_i = \begin{cases} 1 & x_i \; is \; observed \text{ with } p(\delta_i = 1) = p \\ 0 & otherwise \end{cases} \quad (2)$$

See [Zhao (2012), Richard et al. (2019)].

3-Gamma Distribution

The probability density function of gamma distribution has the form:

$$f(x_i;k,\theta) = x_i^{k-1} \frac{e^{-x_i/\theta}}{\theta^k \Gamma k} \quad (3)$$

for $x_i \geq 0$ and $k\ (shape), \theta\ (scale) > 0, i = 1,2,\dots,n$

$$E(x_i) = k\theta, var(x_i) = k\theta^2 \quad (4)$$

$$E(\ln x_i) = \psi(k) + \ln\theta\,, var(\ln x_i) = \psi_1(k) \quad (5)$$

where Ψ(k) is the digamma function:

$$\Psi(k) = d\frac{\ln\Gamma(k)}{dk} = -\gamma - \varsigma(4)\big(-(k-1)\big) - \sum_{r=2}^{\infty} \varsigma(r+3)\big(-(k-1)\big)^r \quad (6)$$

where $\gamma = \lim_{n\to\infty}\left[\left(\sum_{s=1}^{\infty}\frac{1}{s}\right) - \ln(n)\right] = 0.57721$ ... is the Euler-mascheroni constant,

and $\varsigma(s) = \sum_{n=1}^{\infty}(n^{-s})$ is the Riemann zeta function.

and $\psi_1(k)$ is the trigamma function:

$$\psi_1(k) = \frac{d}{dk}\psi(k) \quad (7)$$

See [Giles and feng (2009) and Ke et al. (2023)].

4. Estimation of Parameters

In this section, the maximum likelihood technique is used to produce parameter estimates for the one and two gamma distribution populations when some data are missing.

4.1 For One Population :

The sample is denoted as $(x_1, x_2, \dots, x_n)$ with $\theta, k > 0$ are unknown parameters .

The maximum likelihood method to estimate the parameters of the gamma distribution when data are missing is derived.

The maximum likelihood function has the form:

$$L(\theta,k) = \prod_{i=1}^{n}(f(x_i;\theta,k))^{\delta_i} \quad (8)$$

$$L(\theta,k)=\prod_{i=1}^{n}(x_i^{k-1}\frac{e^{-x_i/\theta}}{\theta^k\Gamma k})^{\delta_i} \quad (9)$$

Hence, the logarithm of the likelihood function is given by

$$\ln L(\theta,k)=\sum_{i=1}^{n}\delta_i\left[(k-1)\ln x_i-\frac{x_i}{\theta}-k\ln\theta-\ln\Gamma k\right] \quad (10)$$

$$\frac{\partial(\ln L(\theta,k))}{\partial\theta}=\sum_{i=1}^{n}\left[\delta_i\frac{x_i}{\theta^2}-\delta_i\frac{k}{\theta}\right] \quad (11)$$

Solving the equation, it is obtained

$$\sum_{i=1}^{n}\left[\delta_i\frac{x_i}{\theta^2}-\delta_i\frac{k}{\theta}\right]=0 \quad (12)$$

$$\sum_{i=1}^{n}\delta_i x_i=\theta k\sum_{i=1}^{n}\delta_i \quad (13)$$

The estimator of parameter θ has the form:

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\delta_i x_i}{\hat{k}\sum_{i=1}^{n}\delta_i} \quad (14)$$

4.1.2 For Parameter *k*:

$$\frac{\partial(\ln L(\theta,k))}{\partial k}=\sum_{i=1}^{n}[\delta_i\ln x_i-\delta_i\ln\theta-\delta_i\Psi(k)] \quad (15)$$

$$\Psi(k)=-\gamma-\varsigma(4)\big(-(k-1)\big)-\sum_{r=2}^{\infty}\varsigma(r+3)\big(-(k-1)\big)^r \quad (16)$$

For simplicity, assume that $\sum_{r=2}^{\infty}\varsigma(r+3)\big(-(k-1)\big)^r$ tends to zero

By substitution in (15):

$$\frac{\partial(\ln \mathrm{L}(\theta,\mathrm{k}))}{\partial \mathrm{k}}=\sum_{i=1}^{n}\left[\delta_i\ln x_i-\delta_i\ln\theta-\delta_i\left(-\gamma-\varsigma(4)\big(-(k-1)\big)\right)\right] \quad (17)$$

$$\frac{\partial(\ln \mathrm{L}(\theta,\mathrm{k}))}{\partial \mathrm{k}}=\sum_{i=1}^{n}\big[\delta_i\ln x_i-\delta_i\ln\theta-\delta_i\big(-\gamma+\varsigma(4)k-\varsigma(4)\big)\big] \quad (18)$$

Solving the equation, it is obtained

$$\sum_{i=1}^{n}\big[\delta_i\ln x_i-\delta_i\ln\theta-\delta_i\big(-\gamma+\varsigma(4)k-\varsigma(4)\big)\big]=0 \quad (19)$$

$$\sum_{i=1}^{n}\delta_i\ln x_i-\ln\theta\sum_{i=1}^{n}\delta_i+\gamma\sum_{i=1}^{n}\delta_i-\varsigma(4)k\sum_{i=1}^{n}\delta_i+\varsigma(4)\sum_{i=1}^{n}\delta_i=0 \quad (20)$$

$$\sum_{i=1}^{n}\delta_i\ln x_i+(\gamma+\varsigma(4)-\ln\theta)\sum_{i=1}^{n}\delta_i-\varsigma(4)k\sum_{i=1}^{n}\delta_i=0 \quad (21)$$

$$\sum_{i=1}^{n}\delta_i\ln x_i+(\gamma+\varsigma(4)-\ln\theta)\sum_{i=1}^{n}\delta_i=\varsigma(4)k\sum_{i=1}^{n}\delta_i \quad (22)$$

$$\sum_{i=1}^{n}\delta_i\ln x_i+Q\sum_{i=1}^{n}\delta_i=\varsigma(4)k\sum_{i=1}^{n}\delta_i \quad (23)$$

where $Q=(\gamma+\varsigma(4)-\ln\theta)$

The estimator of parameter k has the form:

$$\hat{k}=\frac{\sum_{i=1}^{n}\delta_i\ln x_i+\hat{Q}\sum_{i=1}^{n}\delta_i}{\varsigma(4)\sum_{i=1}^{n}\delta_i} \quad (24)$$

where $\hat{Q}=\big(\gamma+\varsigma(4)-\ln\hat{\theta}\big)$

4.2 For Two Populations:

When there are any missing data in the joint likelihood function. The maximum likelihood method is used to estimate the parameters of the Gamma distribution for two independent populations.

4.2.1 For Parameter θ:

Assume that $\theta_1 = \theta_2 = \theta$ where θ is unknown, the likelihood function of θ is:

$$L(\theta, k_1, k_2) = \prod_{i=1}^{n} f(x_i, \theta_1, k_1) \prod_{i=1}^{n} f(y_i, \theta_2, k_2) \quad (25)$$

$$L(\theta, k_1, k_2) = \prod_{i=1}^{n} \left(x_i^{k_1-1} \frac{e^{-x_i/\theta_1}}{\theta_1^{k_1}\Gamma k_1}\right)^{\delta_i} \prod_{i=1}^{n} \left(y_i^{k_2-1} \frac{e^{-y_i/\theta_2}}{\theta_2^{k_2}\Gamma k_2}\right)^{\eta_i} \quad (26)$$

Hence, the logarithm of the likelihood function is given by

$$\ln L(\theta, k_1, k_2) = \sum_{i=1}^{n} \delta_i \left[(k_1 - 1)\ln x_i - \frac{x_i}{\theta} - k_1 \ln\theta - \ln\Gamma k_1\right] + \sum_{i=1}^{n} \eta_i \left[(k_2 - 1)\ln y_i - \frac{y_i}{\theta} - k_2 \ln\theta - \ln\Gamma k_2\right] \quad (27)$$

$$\frac{\partial(\ln L(\theta, k_1, k_2))}{\partial\theta} = \sum_{i=1}^{n} \left[\delta_i \frac{x_i}{\theta^2} - \delta_i \frac{k_1}{\theta}\right] + \sum_{i=1}^{n} \left[\eta_i \frac{y_i}{\theta^2} - \eta_i \frac{k_2}{\theta}\right] \quad (28)$$

Solving the equation, it is obtained

$$\sum_{i=1}^{n} \left[\delta_i \frac{x_i}{\theta^2} - \delta_i \frac{k_1}{\theta}\right] + \sum_{i=1}^{n} \left[\eta_i \frac{y_i}{\theta^2} - \eta_i \frac{k_2}{\theta}\right] = 0 \quad (29)$$

$$\sum_{i=1}^{n} \delta_i \frac{x_i}{\theta^2} - \sum_{i=1}^{n} \delta_i \frac{k_1}{\theta} + \sum_{i=1}^{n} \eta_i \frac{y_i}{\theta^2} - \sum_{i=1}^{n} \eta_i \frac{k_2}{\theta} = 0 \quad (30)$$

$$\sum_{i=1}^{n} \delta_i \frac{x_i}{\theta^2} + \sum_{i=1}^{n} \eta_i \frac{y_i}{\theta^2} = \sum_{i=1}^{n} \delta_i \frac{k_1}{\theta} + \sum_{i=1}^{n} \eta_i \frac{k_2}{\theta} \quad (31)$$

$$\frac{\sum_{i=1}^{n} \delta_i x_i + \sum_{i=1}^{n} \eta_i y_i}{\theta^2} = \frac{k_1 \sum_{i=1}^{n} \delta_i + k_2 \sum_{i=1}^{n} \eta_i}{\theta} \quad (32)$$

$$\sum_{i=1}^{n} \delta_i x_i + \sum_{i=1}^{n} \eta_i y_i = \theta(k_1 \sum_{i=1}^{n} \delta_i + k_2 \sum_{i=1}^{n} \eta_i) \quad (33)$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \delta_i x_i + \sum_{i=1}^{n} \eta_i y_i}{\hat{k}_1 \sum_{i=1}^{n} \delta_i + \hat{k}_2 \sum_{i=1}^{n} \eta_i} \quad (34)$$

4.2.2 For Parameter k:

Assume $k_1 = k_2 = k$ where k is unknown, the observation likelihood function of k is

$$L(\theta_1, \theta_2, k) = \prod_{i=1}^{n} \left(x_i^{k_1-1} \frac{e^{-x_i/\theta_1}}{\theta_1^{k_1}\Gamma k_1}\right)^{\delta_i} \prod_{i=1}^{n} \left(y_i^{k_2-1} \frac{e^{-y_i/\theta_2}}{\theta_2^{k_2}\Gamma k_2}\right)^{\eta_i} \quad (35)$$

Hence, the logarithm of the likelihood function is given by

$$\ln L(\theta_1, \theta_2, k) = \sum_{i=1}^{n} \delta_i \left[(k - 1)\ln x_i - \frac{x_i}{\theta_1} - k \ln\theta_1 - \ln\Gamma k\right] + \sum_{i=1}^{n} \eta_i \left[(k - 1)\ln y_i - \frac{y_i}{\theta_2} - k \ln\theta_2 - \ln\Gamma k\right] \quad (36)$$

$$\frac{\partial(\ln L(\theta_1,\theta_2,k))}{\partial k} = \sum_{i=1}^{n}[\delta_i \ln x_i - \delta_i \ln\ \theta_1\ - \delta_i \Psi(k)] + \sum_{i=1}^{n}[\eta_i \ln y_i - \eta_i \ln \theta_2 - \eta_i \Psi(k)] \quad (37)$$

$$\Psi(k) = -\gamma - \varsigma(4)\big(-(k-1)\big) - \sum_{r=2}^{\infty} \varsigma(r+3)\big(-(k-1)\big)^r \quad (38)$$

For simplicity, assume that $\sum_{r=2}^{\infty} \varsigma(r+3)\big(-(k-1)\big)^r$ tends to zero

So

$$\Psi(k) = (-\gamma - \varsigma(4)\big(-(k-1)\big)) \quad (39)$$

By substitution in (37):

$$\frac{\partial(\ln L(\theta_1,\theta_2,k))}{\partial k} = \sum_{i=1}^{n}\big[\delta_i \ln x_i - \delta_i \ln\ \theta_1\ - \delta_i(-\gamma - \varsigma(4)\big(-(k-1)\big))\big] + \quad (40)$$

$$\sum_{i=1}^{n}\big[\eta_i \ln y_i - \eta_i \ln \theta_2 - \eta_i(-\gamma - \varsigma(4)\big(-(k-1)\big))\big]$$

Solving the equation, it is obtained

$$\sum_{i=1}^{n}[\delta_i \ln x_i - \delta_i \ln \theta_1 - \delta_i\big(-\gamma + \varsigma(4)k - \varsigma(4)\big)] + \sum_{i=1}^{n}[\eta_i \ln y_i - \eta_i \ln \theta_2 - \quad (41)$$

$$\eta_i\big(-\gamma + \varsigma(4)k - \varsigma(4)\big)] = 0$$

$$\sum_{i=1}^{n} \delta_i \ln x_i - \ln \theta_1 \sum_{i=1}^{n} \delta_i + \gamma \sum_{i=1}^{n} \delta_i - \varsigma(4)k \sum_{i=1}^{n} \delta_i + \varsigma(4) \sum_{i=1}^{n} \delta_i \quad (42)$$

$$+ \sum_{i=1}^{n} \eta_i \ln y_i - \ln \theta_2 \sum_{i=1}^{n} \eta_i + \gamma \sum_{i=1}^{n} \eta_i - \varsigma(4)k \sum_{i=1}^{n} \eta_i + \varsigma(4) \sum_{i=1}^{n} \eta_i = 0$$

$$(\sum_{i=1}^{n} \delta_i \ln x_i + \sum_{i=1}^{n} \eta_i \ln\ y_i) - (\ln \theta_1 \sum_{i=1}^{n} \delta_i + \ln \theta_2 \sum_{i=1}^{n} \eta_i) + \gamma(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i) \quad (43)$$

$$. - \varsigma(4)k\left(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i\right) + \varsigma(4)\left(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i\right) = 0$$

$$\varsigma(4)k(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i) = (\sum_{i=1}^{n} \delta_i \ln x_i + \sum_{i=1}^{n} \eta_i \ln y_i) - (\ln \theta_1 \sum_{i=1}^{n} \delta_i + \quad (44)$$

$$\ln \theta_2 \sum_{i=1}^{n} \eta_i) + \gamma(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i) + \varsigma(4)(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i)$$

The estimator of parameter k has the form:

$$\hat{k} = \frac{(\sum_{i=1}^{n} \delta_i \ln\ x_i + \sum_{i=1}^{n} \eta_i \ln y_i) - (\ln \theta_1 \sum_{i=1}^{n} \delta_i + \ln \theta_2 \sum_{i=1}^{n} \eta_i) + \big(\gamma + \varsigma(4)\big)(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i)}{\varsigma(4)(\sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \eta_i)} \quad (45)$$

5. Results and Discussion

The parameters of the gamma distributions were determined by simulation with varied sample sizes (10, 30, and 50) and missing data percentages (10%, 20%, and 30%). The parameter estimator is produced using the maximum likelihood technique for missing data. For one population, equations (14) and (24) are utilised, but equations (34) and (45) are used for two populations. The listwise deletion method and mean imputation are also used to compare the findings. The chain has 10,000 iterations. True parameters ($\theta_1 = 2\,, k_1 = 1$) are utilised for a one

population. Assuming that $\theta_1 = \theta_2 = \theta$and $k_1 = k_2 = k$ the real parameters for the two gamma populations are , $(\theta = 1.5\,, k = 0.75)$. The three techniques' parameter estimators are compared using the mean square error and the relative absolute biases (RAB).

The mean square error and the relative absolute biases are defined as :

MSE = $\mathrm{E}(estimator - true\ value)^2$ (46)

RAB = $\frac{\lfloor estimator - true\ value \rfloor}{true\ value}$ (47)

For one population of the gamma distribution with missing data for various sample sizes, the mean square error and the relative absolute biases of the estimators are shown in Tables (1) and (2), while the results for two populations are shown in Tables (3) and (4).

Table (1) : The Mean Square Error of Estimators for One Population

Percentage of Missing	Sample size	10			30			50		
	Method	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation
10%	$\theta_1 = 2$	0.2239877	0.2286125	0.2235253	0.0721273	0.07631378	0.07423426	0.04563136	0.04400976	0.04365309
	$k_1 = 1$	0.1666223	0.1449773	0.1530339	0.1990319	0.1012322	0.1143055	0.19156925	0.09369703	0.1065473
20%	$\theta_1 = 2$	0.2502455	0.2481877	0.2530886	0.09233464	0.08309125	0.08284736	0.0499313	0.04937415	0.0502388
	$k_1 = 1$	0.1864223	0.1485909	0.1723748	0.1990319	0.1033932	0.1333416	0.19126975	0.09509053	0.1233568
30%	$\theta_1 = 2$	0.2853281	0.2942233	0.2809823	0.09291987	0.09404038	0.09302321	0.05758562	0.05830457	0.057025
	$k_1 = 1$	0.1965233	0.1612376	0.1898331	0.1990319	0.1081247	0.1508841	0.18155925	0.0965087	0.1408961

Table (2) : The Relative Absolute Biases of Estimators for One Population

Percentage of Missing	Sample size	10			30			50		
	Method	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation
10%	$\theta_1 = 2$	0.103034	0.105162	0.102822	0.033179	0.035104	0.034148	0.02099	0.020244	0.02008
	$k_1 = 1$	0.074846	0.06669	0.070396	0.047555	0.046567	0.052581	0.042122	0.043101	0.049012
20%	$\theta_1 = 2$	0.115113	0.114166	0.116421	0.039874	0.038222	0.03811	0.022968	0.022712	0.02311
	$k_1 = 1$	0.069856	0.068352	0.079292	0.052587	0.047561	0.061337	0.047122	0.043742	0.056744
30%	$\theta_1 = 2$	0.131251	0.135343	0.129252	0.042743	0.043259	0.042791	0.026489	0.02682	0.026232
	$k_1 = 1$	0.082846	0.074169	0.087323	0.065355	0.049737	0.069407	0.047152	0.044394	0.064812

Table (3) : The Mean Square Error of Estimators for Two Population

Percentage of Missing	Sample size	10			30			50		
	Method	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation
10%	θ = 1.5	1.659086	1.667111	1.631403	1.459658	1.464337	1.460161	1.432357	1.424886	1.413819
	k =.75	0.6625628	0.565111	0.6137946	0.5348395	0.5287729	0.5808246	0.5362115	0.5203232	0.5738813
20%	θ = 1.5	1.695932	1.700456	1.711938	1.477018	1.480671	1.467662	1.432115	1.42385	1.432312
	k =.75	0.5825638	0.57021	0.6719239	0.5342385	0.5321467	0.6396705	0.616012	0.520483	0.6375328
30%	θ = 1.5	1.748196	1.742285	1.758395	1.497167	1.503361	1.479253	1.43112	1.432137	1.432793
	k =.75	0.7625322	0.5811721	0.727947	0.6247294	0.5347349	0.700992	0.5364115	0.5219587	0.6966006

Table (4) : The Relative Absolute Biases of Estimators for One Population

Percentage of Missing	Sample size	10			30			50		
	Method	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation	Maximum Likelihood	Listwise	Mean Imputation
10%	θ = 1.5	0.76318	0.766871	0.750445	0.671443	0.673595	0.671674	0.658884	0.655448	0.650357
	k =.75	0.258779	0.259951	0.282346	0.241426	0.243236	0.267179	0.237365	0.239349	0.263985
20%	θ = 1.5	0.780129	0.78221	0.787491	0.679428	0.681109	0.675125	0.658773	0.654971	0.658864
	k =.75	0.258779	0.262297	0.309085	0.24575	0.244787	0.294248	0.237365	0.239422	0.293265
30%	θ = 1.5	0.80417	0.801451	0.808862	0.688697	0.691546	0.680456	0.658315	0.658783	0.659085
	k =.75	0.258779	0.267339	0.334856	0.287376	0.245978	0.322456	0.237365	0.240101	0.320436

The results from the tables above can be used to illustrate the following for both the mean square error and the relative absolute biases:

5.1 For Sample Size = 10 .

When 10% of the data is missing, the mean imputation technique's scale parameter (θ) is smaller than the other two techniques' for both the one and the two populations. Nevertheless, the listwise elimination approach is the least efficient for the shape parameter (k).

For all estimator, the mean imputation and maximum likelihood exceed the listwise deletion method. Only in two populations is the maximum likelihood technique smaller for the scale parameter. when the 20% missing data percentage.

The two populations' scale parameter and the one and two populations' shape parameter are the lowest for the listwise deletion method. Conversely, the scale parameter for the single population has the lowest mean imputation value. when the amount of missing data is 30%.

5.2 For Sample Size =30 .

The maximum likelihood method's scale parameter value is at its lowest when the fraction of missing data is 10%. On the other hand, the smallest value for the shape parameter is the listwise deletion strategy.

The scale parameter for one population has the lowest value according to the maximum likelihood technique. The mean imputation value is the lowest for two populations. The listwise elimination method is the least values for the shape parameter. When missing proportion is 20%

For the missing percentage of 30%, The maximum likelihood technique has the lowest value for the scale parameter in one population. For two populations, the imputation mean value is lowest. The listwise deletion approach uses the shape parameter's fewest values.

5.3 For Sample Size = 50 .

For the missing percentage of 10%, the mean imputation approach is less than the other two methods for the scale parameters. While the listwise deletion method has the lowest shape parameter value.

For the20% missing percentage, the listwise deletion approach yields the least values for all the parameters.

For the missing percentage of 30%, One population's scale parameter has the lowest value, based on the mean imputation approach. The maximum likelihood value is lowest for two populations. The listwise elimination method yields the shape parameter's lowest values.

6. Conclusion

The following are the primary conclusions drawn from the simulation study for both the one and two populations:

1- The three approaches (maximum likelihood method, listwise deletion method, and mean imputation) show only slight variations in the mean square error and relative absolute biases.

2- When one is interested in the scale parameter θ, one should apply the mean imputation strategy. Instead, employ the listwise deletion method if you are interested in the shape parameter k.

3. As the amount of missing data increases, it is preferable to use the maximum likelihood approach or mean imputation for the scale parameter θ. However, for the shape parameter k, it is better to employ the listwise deletion method.

4-For the scale parameter θ, the mean imputation strategy is better as sample size grows, but for the shape parameter k, the listwise deletion method is better.

References

1) Acock , A.C. (2005). "Working with Missing Values*". Journal of Marriage and Family, Vol. 67, pp. 1012–1028.*
2) Baraldi, A.N. and Enders, C.K., (2010), *An introduction to modern missing data analyses*, Journal of School Psychology, vol. 48, pp. 5–37.
3) Donders, A.T., Geert J.M.G., Stijnen, T. and Karel G.M., (2006), *Review: A gentle introduction to imputation of missing values*, Journal of Clinical Epidemiology, vol. 59, pp. (1087-1091).
4) Dong Y, Peng CY. , (2013) "*Principled missing data methods for researchers"*. Springerplus. 14;2(1).
5) Farzandi, M. , Sanaeinejad, H., Rezaei-Pazhan, H. and Sarmad, M.(2022). "Improving estimation of missing data in historical monthly precipitation by evolutionary methods in the semi-arid area*". Environment, Development and Sustainability: A Multidisciplinary Approach to the Theory and Practice of Sustainable Development, Vol. 24(6), pp .8313-8332.*
6) Giles, D.E. and Feng, H. (2009). *Bias of the Maximum Likelihood Estimators of the Two-Parameter Gamma Distribution Revisited*, Department of Economics, University of Victoria, university of victoria, pp. 1-19.
7) Golden, R.M.; Henley, S.S.; White, H.; Kashner, T.M.(2019) Consequences of Model Misspecification for Maximum Likelihood Estimation with Missing Data. *Econometrics 7,* No. 41, 37.
8) Gupta, V. K. and Grover, G. (2017). "Multiple imputation for gamma outcome variable using generalized linear model". *Journal of Statistical Computation and Simulation., Vol. 87, Issue 10, pp .1980-1988.*
9) Jäger, S., Allhorn, A. and Biebmann, F. A. (2021)." Benchmark for Data Imputation Methods". *Front Big Data Vol. 4.*

10) Ke, X., Sirao W., Min Z. and Huajun Y.(2023). "New Approaches on Parameter Estimation of the Gamma Distribution" .*Mathematics, 11, no. 4: 927.*

11) Kumar, A., Boehm, M., and Yang, J. (2017). “Data Management in Machine Learning” in Proc. ACM SIGMOD Int. Conf. Manag. Data Part, Chicago Illinois USA (Association for Computing Machinery), 1717–1722.

12) Luo, D., (2013), *Estimation and Test for Two Pareto Populations with Partially Missing Data*, Applied Mathematical Sciences, vol. 7, No. 41, pp. 2027 – 2033.

13) Nguyen, T., Nguyen-Duy, K.M., Nguyen, D.H.M., Nguyen, B.T. and Wade, B.A., 2021. DPER: Efficient Parameter Estimation for Randomly Missing Data. *arXiv preprint arXiv:2106.05190*.

14) Pepinsky, T. B. (2018). “A note on listwise deletion versus multiple imputation”. *Political Analysis. Vol. 26, Issue 4, pp 480–488.*

15) Richard M. Golden & Steven S. Henley & Halbert White & T. Michael Kashner, 2019. "Consequences of Model Misspecification for Maximum Likelihood Estimation with Missing Data," *Econometrics, MDPI, vol. 7(3), pp 1-27.*

16) Zhao, Z., Wang, S., Wang, R. and Liling, (2009), *Parameter Estimation and Hypothesis Testing of Two Exponential Populations under Type I Censoring Sample with Missed Data*, Journal of Jilin University (Science Edition), vol. 47, No. 1, pp. 26-30.

17) Zhao, Z., (2012), *Parameter Estimation and Hypothesis Testing of Two Negative Binomial Distribution Population with Missing Data*, International Conference on Medical Physics and Biomedical Engineering, vol. 33, pp. 1475 – 1480.