STATISTICS IN TRANSITION new series, December 2020
Vol. 21, No. 5, pp. 41–60, DOI 10.21307/stattrans-2020-054
Received – 12.10.2019; accepted – 04.05.2020

Comparing particulate matter dispersion in Thailand using the Bayesian Confidence Intervals for ratio of coefficients of variation

Warisa Thangjai[1], Suparat Niwitpong[2]

ABSTRACT

Recently, harmful levels of air pollution have been detected in many provinces of Thailand. Particulate matter (PM) contains microscopic solids or liquid droplets that are so small that they can be inhaled and cause serious health problems. A high dispersion of PM is measured by a coefficient of variation of log-normal distribution. Since the log-normal distribution is often used to analyse environmental data such as hazardous dust particle levels and daily rainfall data. These data focus the statistical inference on the coefficient of variation. In this paper, we develop confidence interval estimation for the ratio of coefficients of variation of two log-normal distributions constructed using the Bayesian approach. These confidence intervals were then compared with the existing approaches: method of variance estimates recovery (MOVER), modified MOVER, and approximate fiducial approaches using their coverage probabilities and average lengths via Monte Carlo simulation. The simulation results show that the Bayesian confidence interval performed better than the others in terms of coverage probability and average length. The proposed approach and the existing approaches are illustrated using examples from data set PM10 level and PM2.5 level in the northern Thailand.

Key words: Bayesian approach, coefficient of variation, confidence interval, log-normal distribution, ratio.

1. Introduction

Nowadays, the problem of air pollution has received widespread attention in toxicology and epidemiology studies because it is associated with increased incidences of human disease and mortality rate (Xing et al., 2016). The effects on human health include the cardiovascular system, resulting in heart attacks and heart failure, and the respiratory tract, resulting in asthma and bronchitis. Smoke, dust, and smog create air pollution, which includes gaseous pollutants and particulate matter (PM): the gases include carbon monoxide, sulphur dioxide, ozone, and nitrogen dioxide, while PM is defined by size, e.g. PM2.5 ($\leq 2.5\ \mu m$) and PM10 ($\leq 10\ \mu m$), and so on. People are at high risk when they live in high PM levels. For PM2.5, both short-term and long-term exposure has been associated with increased hospital admission and absenteeism from school, work, etc. Exposure to

[1]Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok, 10240, Thailand. E-mail: wthangjai@yahoo.com. ORCID: https://orcid.org/0000-0002-9306-3742.

[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand. E-mail: suparat.n@sci.kmutnb.ac.th. ORCID: https://orcid.org/0000-0003-3059-1131.

PM2.5 can also result in emergency room visits for asthma symptoms whereas exposure to the PM10 can result in hospitalization of chronic lung disease and/or premature death. Moreover, PM2.5 and PM10 can damage stone and culturally important objects such as monuments and statues. Thailand is a country located in Southeast Asia. It covers a total land area of approximately 513,000 km^2 and is divided into six regions used in geographic studies: north, northeast, central, east, west, and south. These are based on natural features and human cultural patterns. Recently, Thailand has faced the PM problem resulting in the deterioration of air quality. Harmful levels have been detected in the north region of Thailand, in Chiang Mai, Chiang Rai, Lampang, Mae Hong Son, Nan, Phrae, and Phayao provinces. The coefficient of variation can be used as a statistic to describe air quality and thus can be used to measure and manage air pollution risk.

Meanwhile, several authors have discussed which parameter should be used in statistical inference for a log-normal distribution (Lacey et al., 1997; Royston, 2001; Krishnamoorthy and Mathew, 2003; Hannig et al., 2006; Tian and Wu, 2007; Sharma and Singh, 2010; Harvey and van der Merwe, 2012; Lin and Wang, 2013; Rao and D'Cunha, 2016; Thangjai et al., 2016; Nam and Kwon, 2017; Hasan and Krishnamoorthy, 2017; Thangjai and Niwitpong, 2019). Furthermore, the coefficient of variation has been used in various applications (Tsim et al., 1991; Faupel-Badger et al., 2010). In addition, the confidence intervals for the coefficient of variation have received some attention recently (Niwitpong, 2013; Ng, 2014; Thangjai et al., 2016; Nam and Kwon, 2017; Hasan and Krishnamoorthy, 2017). The inference with the log-normal coefficient of variation is interesting. Nam and Kwon (2017) proposed the method of variance estimate recovery (MOVER) approach for constructing the confidence intervals for the ratio of coefficients of variation of log-normal distributions. Meanwhile, Hasan and Krishnamoorthy (2017) improved the confidence intervals for the ratio of coefficients of variation of log-normal distributions based on an alternative MOVER approach and the fiducial approach.

Both these approaches have produced classical statistics, and while some problems are best solved using these, others are best solved using the Bayesian approach. Therefore, in this paper, we extend the research idea from Hasan and Krishnamoorthy (2017) to develop the Bayesian approach for confidence interval estimation of the ratio of coefficients of variation of log-normal distributions. The Bayesian approach is a statistical method based on Bayes' theorem, which is used to update the probability. The method derives the posterior probability that is the result of a prior probability and a likelihood function. This is advantageous in the interpretation and construction of the Bayesian confidence interval, which makes it more straightforward than the classical confidence interval approaches. However, a disadvantage is that the Bayesian confidence interval requires more input than the classical approach (Casella and Berger, 2002). The Bayesian approach for parameter estimation has been addressed in several research papers (Harvey and van der Merwe, 2012; Rao and D'Cunha, 2016; Ma and Chen, 2018).

2. Methods

Suppose that random samples X_1 and X_2 follow two independent normal distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Also, suppose that Y_1 and Y_2

are random samples of sizes n_1 and n_2 from two independent log-normal distributions with parameters μ_1, σ_1^2, μ_2, and σ_1^2, respectively. The mean and variance of Y_1 are

$$E(Y_1) = \exp(\mu_1 + \sigma_1^2/2) \quad \text{and} \quad Var(Y_1) = (\exp(\sigma_1^2) - 1)(\exp(2\mu_1 + \sigma_1^2)). \tag{1}$$

The coefficient of variation of Y_1 is

$$\tau_1 = E(Y_1)/\sqrt{Var(Y_1)} = \sqrt{\exp(\sigma_1^2) - 1}. \tag{2}$$

Similarly, the mean and variance of Y_2 are

$$E(Y_2) = \exp(\mu_2 + \sigma_2^2/2) \quad \text{and} \quad Var(Y_2) = (\exp(\sigma_2^2) - 1)(\exp(2\mu_2 + \sigma_2^2)). \tag{3}$$

The coefficient of variation of Y_2 is

$$\tau_2 = E(Y_2)/\sqrt{Var(Y_2)} = \sqrt{\exp(\sigma_2^2) - 1}. \tag{4}$$

The ratio of two coefficients of variation is given by

$$\theta = \frac{\tau_1}{\tau_2} = \sqrt{\frac{\exp(\sigma_1^2) - 1}{\exp(\sigma_2^2) - 1}}. \tag{5}$$

The estimator of θ is

$$\hat{\theta} = \frac{\hat{\tau}_1}{\hat{\tau}_2} = \sqrt{\frac{\exp(S_1^2) - 1}{\exp(S_2^2) - 1}}, \tag{6}$$

where S_1^2 and S_2^2 are the variances of the log-transformed sample from a log-normal distributions.

This section describes the three existing confidence intervals. One is the MOVER confidence interval introduced by Nam and Kwon (2017). The modified MOVER and approximate fiducial confidence intervals are proposed by Hasan and Krishnamoorthy (2017). Furthermore, the Bayesian confidence interval, which is a novel approach, is presented.

2.1. Classical confidence intervals for ratio of coefficients of variation

Three confidence intervals for the ratio of coefficients of variation of log-normal distributions are presented.

2.1.1 MOVER confidence interval for ratio of coefficients of variation

Donner and Zou (2002) and Zou and Donner (2008) describe a theorem of MOVER. The lower limit L and the upper limit U are used to derive the variance estimates for θ, which is ranging from L to $\hat{\theta}$ and from $\hat{\theta}$ to U. The variance estimate recovered from the lower tail of θ is $(\hat{\theta} - L)^2/z^2$, where z denotes the $100(\alpha/2)$-th percentile of the standard normal distribution. Similarly, the variance estimate recovered from the upper tail of θ is

$(U-\hat{\theta})^2/z^2$. These variance estimates are used to construct the lower and upper limits of the confidence interval for θ.

Nam and Kwon (2017) introduced the MOVER approach for constructing the confidence interval for the ratio of coefficients of variation of two log-normal distributions. The MOVER confidence interval can be obtained from the one for $\ln(\theta)=\ln(\tau_1)-\ln(\tau_2)$. The variances of $\ln(\hat{\tau}_1)$ and $\ln(\hat{\tau}_2)$ are given by

$$\hat{Var}(\ln(\hat{\tau}_1))=\frac{\hat{\sigma}_1^2(1+\hat{\tau}_1^2)^2}{2n_1\hat{\tau}_1^4} \tag{7}$$

and

$$\hat{Var}(\ln(\hat{\tau}_2))=\frac{\hat{\sigma}_2^2(1+\hat{\tau}_2^2)^2}{2n_2\hat{\tau}_2^4}, \tag{8}$$

where $\hat{\sigma}_1^2=(n_1-1)S_1^2/n_1$ and $\hat{\sigma}_2^2=(n_2-1)S_2^2/n_2$ are the maximum likelihood estimates of σ_1^2 and σ_2^2, respectively.

The confidence intervals of $\ln(\tau_1)$ and $\ln(\tau_2)$ are given by

$$[l_1^{'},u_1^{'}]=[\ln(\hat{\tau}_1)-z_{1-\alpha/2}\sqrt{\hat{Var}(\ln(\hat{\tau}_1))},\ln(\hat{\tau}_1)+z_{1-\alpha/2}\sqrt{\hat{Var}(\ln(\hat{\tau}_1))}] \tag{9}$$

and

$$[l_2^{'},u_2^{'}]=[\ln(\hat{\tau}_2)-z_{1-\alpha/2}\sqrt{\hat{Var}(\ln(\hat{\tau}_2))},\ln(\hat{\tau}_2)+z_{1-\alpha/2}\sqrt{\hat{Var}(\ln(\hat{\tau}_2))}], \tag{10}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$-th percentile of the standard normal distribution and $\hat{Var}(\ln(\hat{\tau}_1))$ and $\hat{Var}(\ln(\hat{\tau}_2))$ are defined in Equation (7) and Equation (8).

The lower and upper limits of the confidence interval for $\ln(\theta)=\ln(\tau_1)-\ln(\tau_2)$ based on the MOVER approach are given by

$$L_{\theta.MOVER}=\ln(\hat{\tau}_1)-\ln(\hat{\tau}_2)-\sqrt{(\ln(\hat{\tau}_1)-l_1^{'})^2+(\ln(\hat{\tau}_2)-u_2^{'})^2} \tag{11}$$

and

$$U_{\theta.MOVER}=\ln(\hat{\tau}_1)-\ln(\hat{\tau}_2)+\sqrt{(\ln(\hat{\tau}_1)-u_1^{'})^2+(\ln(\hat{\tau}_2)-l_2^{'})^2}. \tag{12}$$

Therefore, the $100(1-\alpha)\%$ MOVER confidence interval for ratio of coefficients of variation θ is defined as

$$CI_{\theta.MOVER}=[L_{\theta.MOVER},U_{\theta.MOVER}]=[\exp(L_{\theta.MOVER}),\exp(U_{\theta.MOVER})]. \tag{13}$$

2.1.2 Modified MOVER confidence interval for ratio of coefficients of variation

Hasan and Krishnamoorthy (2017) extended the research paper from Nam and Kwon (2017) to propose the new confidence interval for the ratio of coefficients of variation based on the MOVER approach. The new confidence interval is called modified MOVER confidence interval. Hasan and Krishnamoorthy (2017) used the exact confidence intervals for

τ_1^2 and τ_2^2 given by

$$[l_1^{''}, u_1^{''}] = [\exp\left(\frac{(n_1-1)S_1^2}{\chi^2_{n_1-1,\alpha/2}}\right) - 1, \exp\left(\frac{(n_1-1)S_1^2}{\chi^2_{n_1-1,1-\alpha/2}}\right) - 1] \tag{14}$$

and

$$[l_2^{''}, u_2^{''}] = [\exp\left(\frac{(n_2-1)S_2^2}{\chi^2_{n_2-1,\alpha/2}}\right) - 1, \exp\left(\frac{(n_2-1)S_2^2}{\chi^2_{n_2-1,1-\alpha/2}}\right) - 1], \tag{15}$$

where $\chi^2_{n_i-1,1-\alpha/2}$ and $\chi^2_{n_i-1,\alpha/2}$ denote the $100(1-\alpha/2)$-th and $100(\alpha/2)$-th percentiles of the chi-squared distribution with $n_i - 1$ degrees of freedom for $i = 1, 2$.

The lower and upper limits of the modified MOVER confidence interval for $\ln(\tau_1/\tau_2)^2$ are given by

$$L_{MMOVER} = \ln(\hat{\tau}_1^2) - \ln(\hat{\tau}_2^2) - \sqrt{(\ln(\hat{\tau}_1^2) - \ln(l_1^{''}))^2 + (\ln(\hat{\tau}_2^2) - \ln(u_2^{''}))^2} \tag{16}$$

and

$$U_{MMOVER} = \ln(\hat{\tau}_1^2) - \ln(\hat{\tau}_2^2) + \sqrt{(\ln(\hat{\tau}_1^2) - \ln(u_1^{''}))^2 + (\ln(\hat{\tau}_2^2) - \ln(l_2^{''}))^2}, \tag{17}$$

where $\hat{\tau}_1^2 = \exp(S_1^2) - 1$ and $\hat{\tau}_2^2 = \exp(S_2^2) - 1$.

Therefore, the $100(1-\alpha)\%$ modified MOVER confidence interval for ratio of coefficients of variation θ is defined as

$$CI_{\theta.MMOVER} = [L_{\theta.MMOVER}, U_{\theta.MMOVER}] = [\sqrt{\exp(L_{MMOVER})}, \sqrt{\exp(U_{MMOVER})}]. \tag{18}$$

2.1.3 Approximate fiducial confidence interval for ratio of coefficients of variation

The fiducial confidence interval is computed based on a fiducial quantity. The coefficient of variation of log-normal distribution is used the fiducial quantity for σ^2 only. This is because the coefficient of variation is the function of σ^2 only. The percentiles of fiducial generalized pivotal quantity for ratio of coefficients of variation is estimated using simulation. To avoid using the simulation, Hasan and Krishnamoorthy (2017) used modified normal based approximation to construct the approximate fiducial confidence interval. Let s_1^2 and s_2^2 be observed values of S_1^2 and S_2^2, respectively.

The lower and upper limits of the approximate fiducial confidence interval for $\ln(\tau_1/\tau_2)^2$ are given by

$$L_{AF} = \ln(T_{1;0.5}) - \ln(T_{2;0.5}) - \sqrt{(\ln(T_{1;0.5}) - \ln(T_{1;\alpha/2}))^2 + (\ln(T_{2;0.5}) - \ln(T_{2;1-\alpha/2}))^2} \tag{19}$$

and

$$U_{AF} = \ln(T_{1;0.5}) - \ln(T_{2;0.5}) + \sqrt{(\ln(T_{1;0.5}) - \ln(T_{1;1-\alpha/2}))^2 + (\ln(T_{2;0.5}) - \ln(T_{2;\alpha/2}))^2}, \tag{20}$$

where $T_{i;p} = \exp((n_i-1)S_i^2/\chi^2_{n_i-1,p}) - 1$ and $\chi^2_{n_i-1,p}$ is the $100(p)$-th percentile of the chi-squared distribution with $n_i - 1$ degrees of freedom, respectively.

Therefore, the $100(1-\alpha)\%$ approximate fiducial confidence interval for the ratio of coefficients of variation θ is defined as

$$CI_{\theta.AF} = [L_{\theta.AF}, U_{\theta.AF}] = [\sqrt{\exp(L_{AF})}, \sqrt{\exp(U_{AF})}]. \tag{21}$$

2.2. Bayesian confidence interval for ratio of coefficients of variation

Bayesian confidence interval is constructed using the concept of Bayesian inference. The Bayesian confidence interval uses a prior distribution. This distribution is based on the experimenter's belief and is updated with the sample information. The Bayesian confidence interval derives a posterior probability as a consequence of a prior probability and a likelihood function. Posterior probability is computed by Bayes' theorem. Let $X_1 = \ln(Y_1)$ be the normal distribution with mean μ_1 and variance σ_1^2. Also, let $X_2 = \ln(Y_2)$ be the normal distribution with mean μ_2 and variance σ_2^2.The likelihood function for μ_1, μ_2, σ_1^2 and σ_2^2 is

$$\begin{aligned} L(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|data) &\propto \left(\frac{1}{\sigma_1^2}\right)^{n_1/2} \exp\left(-\frac{(n_1-1)s_1^2+n_1(\mu_1-\bar{x}_1)^2}{2\sigma_1^2}\right) \\ &\times \left(\frac{1}{\sigma_2^2}\right)^{n_2/2} \exp\left(-\frac{(n_2-1)s_2^2+n_2(\mu_2-\bar{x}_2)^2}{2\sigma_2^2}\right), \end{aligned} \tag{22}$$

where $i = 1,2$ and $\bar{x}_i$ and s_i^2 are the observed values of $\bar{X}_i$ and S_i^2, respectively.

Taking the logarithm of the likelihood function, the log-likelihood function is obtained by

$$\begin{aligned} \ln(L) &= -\frac{n_1}{2}\ln(\sigma_1^2) - \frac{(n_1-1)s_1^2+n_1(\mu_1-\bar{x}_1)^2}{2\sigma_1^2} \\ &- \frac{n_2}{2}\ln(\sigma_2^2) - \frac{(n_2-1)s_2^2+n_2(\mu_2-\bar{x}_2)^2}{2\sigma_2^2}. \end{aligned} \tag{23}$$

The second derivatives of log-likelihood function with respect to each parameter are

$$\frac{\partial^2 \ln(L)}{\partial\mu_1^2} = -\frac{n_1}{\sigma_1^2} \quad \text{and} \quad \frac{\partial^2 \ln(L)}{\partial\mu_2^2} = -\frac{n_2}{\sigma_2^2}, \tag{24}$$

$$\frac{\partial^2 \ln(L)}{\partial\mu_1\partial\sigma_1^2} = \frac{n_1(\mu_1-\bar{x}_1)}{(\sigma_1^2)^2} \quad \text{and} \quad \frac{\partial^2 \ln(L)}{\partial\mu_2\partial\sigma_2^2} = \frac{n_2(\mu_2-\bar{x}_2)}{(\sigma_2^2)^2}, \tag{25}$$

$$\frac{\partial^2 \ln(L)}{(\partial\sigma_1^2)^2} = \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 - \left(\frac{1}{\sigma_1^2}\right)^3((n_1-1)s_1^2+n_1(\mu_1-\bar{x}_1)^2), \tag{26}$$

$$\frac{\partial^2 \ln(L)}{(\partial\sigma_2^2)^2} = \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 - \left(\frac{1}{\sigma_2^2}\right)^3((n_2-1)s_2^2+n_2(\mu_2-\bar{x}_2)^2), \tag{27}$$

and

$$\frac{\partial^2 \ln(L)}{\partial\sigma_1^2\partial\sigma_2^2} = 0 \quad \text{and} \quad \frac{\partial^2 \ln(L)}{\partial\sigma_2^2\partial\sigma_1^2} = 0. \tag{28}$$

The Fisher information matrix is

$$F(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) = \begin{bmatrix} \frac{n_1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{n_2}{\sigma_2^2} & 0 & 0 \\ 0 & 0 & \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 & 0 \\ 0 & 0 & 0 & \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 \end{bmatrix}. \quad (29)$$

The Bayesian confidence intervals can be construct based on different choices of prior distributions. This paper is interested in the Jeffreys Independence prior. This prior follows from the Fisher information matrix. According the Fisher information matrix, the Jeffreys Independence prior is

$$p(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) = p(\mu_1,\mu_2)p(\sigma_1^2,\sigma_2^2). \quad (30)$$

The joint prior for the mean is

$$p(\mu_1,\mu_2) \propto \begin{vmatrix} \frac{n_1}{\sigma_1^2} & 0 \\ 0 & \frac{n_2}{\sigma_2^2} \end{vmatrix}^{1/2}. \quad (31)$$

The joint prior for the variance is

$$p(\sigma_1^2,\sigma_2^2) \propto \begin{vmatrix} \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 & 0 \\ 0 & \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 \end{vmatrix}^{1/2}. \quad (32)$$

Therefore, the Jeffreys Independence prior is obtained by

$$p(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) \propto \frac{1}{\sigma_1^2}\left(\frac{1}{\sigma_2^2}\right). \quad (33)$$

The conditional posterior distributions of μ_1 and μ_2 are normal distributions. The conditional posterior distributions are given by

$$\mu_1|\sigma_1^2,x_1 \sim N\left(\hat{\mu}_1,\frac{\sigma_1^2}{n_1}\right) \quad (34)$$

and

$$\mu_2|\sigma_2^2,x_2 \sim N\left(\hat{\mu}_2,\frac{\sigma_2^2}{n_2}\right). \quad (35)$$

For σ_1^2 and σ_2^2, the posterior distributions are the inverse gamma distributions given by

$$\sigma_1^2|x_1 \sim IG(\frac{n_1-1}{2},\frac{(n_1-1)s_1^2}{2}) \quad (36)$$

and

$$\sigma_2^2|x_2 \sim IG(\frac{n_2-1}{2}, \frac{(n_2-1)s_2^2}{2}). \tag{37}$$

The posterior distribution of $\ln(\tau_1/\tau_2)^2$ is given by

$$\ln(\theta)^2 = \ln\left(\frac{\tau_1}{\tau_2}\right)^2 = \ln(\exp(\sigma_1^2)-1) - \ln(\exp(\sigma_2^2)-1), \tag{38}$$

where σ_1^2 and σ_2^2 are defined in Equation (36) and Equation (37), respectively.

Let L_{BS} and U_{BS} be the lower and upper limits of the shortest $100(1-\alpha)\%$ highest posterior density interval of $\ln(\theta)^2$, respectively. Therefore, the $100(1-\alpha)\%$ Bayesian confidence interval for ratio of coefficients of variation θ is defined as

$$CI_{\theta.BS} = [L_{\theta.BS}, U_{\theta.BS}] = [\sqrt{\exp(L_{BS})}, \sqrt{\exp(U_{BS})}]. \tag{39}$$

Algorithm 1

Step 1: Generate $\sigma_i^2|x_i \sim IG(\frac{n_i-1}{2}, \frac{(n_i-1)s_i^2}{2})$, where $i = 1,2$.
Step 2: Calculate the value of $\ln(\theta)^2$ as given in Equation (38).
Step 3: Repeat the step 1 - step 2 for q times.
Step 4: Calculate L_{BS} and U_{BS}.
Step 5: Calculate $L_{\theta.BS}$ and $U_{\theta.BS}$.

Algorithm 2

For a given n_1, n_2, μ_1, μ_2, σ_1, σ_2, and θ.
Step 1: Generate x_1 from $N(\mu_1, \sigma_1^2)$ and generate x_2 from $N(\mu_2, \sigma_2^2)$.
Step 2: Calculate $\bar{x}_1$, $\bar{x}_2$, s_1^2 and s_2^2.
Step 3: Construct $CI_{\theta.MOVER(h)} = [L_{\theta.MOVER(h)}, U_{\theta.MOVER(h)}]$.
Step 4: Construct $CI_{\theta.MMOVER(h)} = [L_{\theta.MMOVER(h)}, U_{\theta.MMOVER(h)}]$.
Step 5: Construct $CI_{\theta.AF(h)} = [L_{\theta.AF(h)}, U_{\theta.AF(h)}]$.
Step 6: Construct $CI_{\theta.BS(h)} = [L_{\theta.BS(h)}, U_{\theta.BS(h)}]$.
Step 7: If $L_{(h)} \leq \theta \leq U_{(h)}$ set $p_{(h)} = 1$, else $p_{(h)} = 0$.
Step 8: Calculate $U_{(h)} - L_{(h)}$.
Step 9: Repeat the step 1 - step 8 for a large number of times (say, M times) and calculate coverage probability and average length.

3. Results

The MOVER, modified MOVER, approximate fiducial and Bayesian confidence intervals for ratio of coefficients of variation were conducted to compare the performance. The confidence intervals with the coverage probability greater than or equal to the nominal confidence level of 0.95 and the shortest average length were considered to be the best-performing ones.

Since the log-normal coefficient of variation depends on parameter σ^2 and does not depend on parameter μ, the population means $\mu_1 = \mu_2 = 1$, the population standard deviations (σ_1, σ_2) and sample sizes (n_1, n_2) were varied based on Hasan and Krishnamoorthy's (2017) approach. The coverage probabilities and average lengths were estimated for some assumed values of parameters (σ_1, σ_2) and sample sizes varying from small to moderate. 10,000 random samples were generated using Algorithm 2 for each set of parameters. For the Bayesian confidence interval, 2,500ln$(\theta)^2$'s were obtained by applying Algorithm 1 for each of the random samples.

The coverage probabilities and average lengths of the four confidence intervals are given in Tables 1 and 2. The MOVER confidence intervals attained coverage probabilities under the nominal confidence level of 0.95 for all sample sizes. Meanwhile, the coverage probabilities of the modified MOVER and approximate fiducial confidence intervals were close to the nominal confidence level of 0.95, but their average lengths were not balanced. The Bayesian confidence intervals provided the best coverage probabilities for all sample sizes and the average lengths were shorter than those of the modified MOVER and approximate fiducial confidence intervals. Overall, the Bayesian confidence intervals are preferable in terms of coverage probability and average length.

Table 1: The CP and AL of 95% two-sided confidence intervals for ratio of coefficients of variation of log-normal distributions as function of parameters

(n_1,n_2)	**Approach**	(σ_1,σ_2)							
		(0.1,0.3)		**(0.3,0.7)**		**(0.4,1.2)**		**(0.5,1.6)**	
		CP	**AL**	**CP**	**AL**	**CP**	**AL**	**CP**	**AL**
(4,8)	MOVER	0.8639	0.6160	0.8793	0.8242	0.9031	0.6675	0.9208	0.6745
	MMOVER	**0.9546**	1.2912	**0.9553**	2.8125	**0.9589**	4.4569	**0.9527**	104.7483
	AF	**0.9501**	1.2483	**0.9501**	2.6929	**0.9519**	4.1770	0.9487	93.2188
	BS	0.9485	1.1105	**0.9629**	1.9912	**0.9768**	2.3785	**0.9777**	5.6386
(5,10)	MOVER	0.8882	0.5420	0.9000	0.7301	0.9227	0.5793	0.9243	0.5805
	MMOVER	**0.9563**	0.9144	**0.9567**	1.4768	**0.9544**	1.3068	**0.9502**	1.6590
	AF	**0.9532**	0.8914	**0.9525**	1.4319	**0.9512**	1.2577	0.9470	1.5768
	BS	0.9499	0.8222	**0.9597**	1.2783	**0.9693**	1.1615	**0.9715**	1.3628
(10,5)	MOVER	0.8848	0.7279	0.8823	1.0082	0.8455	0.8969	0.8156	1.0019
	MMOVER	**0.9554**	0.7778	**0.9513**	1.0888	0.9472	0.8510	0.9465	0.7785
	AF	**0.9511**	0.7602	0.9485	1.0625	0.9458	0.8302	0.9457	0.7574
	BS	**0.9558**	0.8067	**0.9618**	1.2256	**0.9621**	1.0964	**0.9613**	1.1222
(7,12)	MOVER	0.9080	0.4707	0.9212	0.6313	0.9326	0.5058	0.9315	0.4974
	MMOVER	**0.9547**	0.6445	**0.9590**	0.9189	**0.9517**	0.7179	**0.9510**	0.6634
	AF	**0.9519**	0.6324	**0.9562**	0.8988	0.9489	0.7012	0.9493	0.6467
	BS	0.9494	0.6055	**0.9597**	0.8679	**0.9663**	0.7264	**0.9699**	0.7154
(10,10)	MOVER	0.9226	0.4725	0.9276	0.6410	0.9234	0.5410	0.9030	0.5532
	MMOVER	**0.9561**	0.5521	**0.9516**	0.7651	0.9465	0.6012	0.9474	0.5374
	AF	**0.9532**	0.5427	0.9485	0.7510	0.9452	0.5911	0.9471	0.5289
	BS	**0.9510**	0.5397	**0.9586**	0.7695	**0.9626**	0.6576	**0.9614**	0.6385

Table 1: Continued

(n_1, n_2)	**Approach**	(σ_1, σ_2)							
		(0.1,0.3)		**(0.3,0.7)**		**(0.4,1.2)**		**(0.5,1.6)**	
		CP	**AL**	**CP**	**AL**	**CP**	**AL**	**CP**	**AL**
(10,15)	MOVER	0.9222	0.4034	0.9285	0.5368	0.9332	0.4347	0.9308	0.4230
	MMOVER	**0.9527**	0.4884	**0.9504**	0.6664	**0.9506**	0.5162	0.9470	0.4560
	AF	0.9494	0.4817	0.9488	0.6561	0.9486	0.5088	0.9458	0.4503
	BS	0.9464	0.4696	**0.9532**	0.6488	**0.9610**	0.5320	**0.9598**	0.5002
(20,20)	MOVER	0.9358	0.3105	0.9438	0.4202	0.9394	0.3451	0.9252	0.3364
	MMOVER	**0.9517**	0.3327	**0.9551**	0.4526	**0.9536**	0.3558	0.9494	0.3147
	AF	0.9496	0.3301	**0.9534**	0.4489	**0.9531**	0.3537	0.9493	0.3138
	BS	0.9489	0.3278	**0.9561**	0.4511	**0.9571**	0.3694	**0.9594**	0.3432

Table 2: The CP AL of 95% two-sided confidence intervals for ratio of coefficients of variation of log-normal distributions as function of sample sizes

(σ_1,σ_2)	**Approach**	(n_1,n_2)							
		(5,5)		**(10,10)**		**(10,20)**		**(30,40)**	
		CP	**AL**	**CP**	**AL**	**CP**	**AL**	**CP**	**AL**
(0.1,0.7)	MOVER	0.8914	0.3437	0.9202	0.2062	0.9257	0.1552	0.9443	0.0980
	MMOVER	**0.9526**	0.4948	0.9466	0.2367	**0.9508**	0.1885	**0.9504**	0.1027
	AF	0.9476	0.4740	0.9442	0.2329	0.9491	0.1863	0.9491	0.1022
	BS	**0.9638**	0.5392	0.9496	0.2406	0.9487	0.1832	0.9484	0.1018
(0.3,0.9)	MOVER	0.8938	0.8596	0.9275	0.5055	0.9283	0.3755	0.9470	0.2370
	MMOVER	**0.9513**	1.4219	**0.9541**	0.5868	0.9490	0.4683	**0.9510**	0.2487
	AF	0.9466	1.3475	**0.9517**	0.5767	0.9458	0.4622	**0.9501**	0.2476
	BS	**0.9666**	1.5980	**0.9609**	0.6099	**0.9513**	0.4566	**0.9520**	0.2478
(0.5,1.4)	MOVER	0.8833	1.0967	0.9156	0.6134	0.9451	0.4188	0.9467	0.2655
	MMOVER	**0.9504**	2.6600	0.9461	0.6597	**0.9571**	0.5162	**0.9550**	0.2699
	AF	0.9456	2.4364	0.9445	0.6478	**0.9544**	0.5093	**0.9546**	0.2691
	BS	**0.9724**	3.2572	**0.9633**	0.7464	**0.9649**	0.5235	**0.9575**	0.2757
(0.3,1.5)	MOVER	0.8650	0.5899	0.9014	0.3268	0.9422	0.2166	0.9425	0.1380
	MMOVER	0.9473	0.7247	0.9499	0.3076	**0.9526**	0.2359	0.9486	0.1356
	AF	0.9441	0.6858	0.9490	0.3043	**0.9513**	0.2340	0.9488	0.1355
	BS	**0.9678**	1.0401	**0.9596**	0.3623	**0.9624**	0.2500	**0.9537**	0.1401
(0.9,1.8)	MOVER	0.8597	2.4916	0.8984	1.1677	0.9478	0.7360	0.9428	0.4259
	MMOVER	**0.9556**	48337.6800	**0.9511**	1.6359	**0.9533**	1.3112	0.9494	0.4250
	AF	**0.9512**	37293.2300	**0.9500**	1.5698	**0.9513**	1.2785	0.9489	0.4235
	BS	**0.9805**	5024.7520	**0.9671**	1.8142	**0.9685**	1.1855	**0.9557**	0.4429
(0.4,0.5)	MOVER	0.8929	2.0732	0.9218	1.2401	0.9195	0.9674	0.9427	0.5952
	MMOVER	**0.9575**	4.7748	**0.9510**	1.5650	0.9481	1.3012	**0.9501**	0.6378
	AF	**0.9524**	4.5000	0.9485	1.5325	0.9462	1.2837	0.9494	0.6344
	BS	**0.9669**	4.0819	**0.9527**	1.5125	0.9489	1.2134	0.9483	0.6259

4. Empirical application

PM10 and PM2.5 from haze smog in Chiang Mai and Nan provinces, in the northern of Thailand, have become serious problems with air pollution having serious effects on health and visibility for transportation. The data in Tables 3 and 6 from the Pollution Control Department show PM10 and PM2.5 levels in Chiang Mai and Nan provinces from 24 March 2019 to 17 April 2019. Moreover, PM2.5 levels in Bangkok and Chiang Rai provinces from 24 March 2019 to 17 April 2019 are presented in Table 9. The confidence intervals for the ratio of coefficients of variation were constructed using these data.

4.1. Example 1

Using Table 3, the statistics of PM10 pollution are summarized in Table 4. In Table 5, the Akaike Information Criterion values support that the two datasets follow log-normal distributions. These two districts were compared with respect to the coefficient of variation. The 95% two-sided confidence intervals were constructed based on the MOVER, modified MOVER, and approximate fiducial approaches, and then compared with the Bayesian approach.

The ratio of the log-normal coefficients of variation for the Chiang Mai and Nan was $\hat{\theta} = 0.9066$. The confidence intervals based on the MOVER, modified MOVER, and approximate fiducial approaches were $CI_{\theta.MOVER} = [0.6009, 1.3676]$ with an interval length of 0.7667, $CI_{\theta.MMOVER} = [0.5829, 1.3977]$ with an interval length of 0.8148, and $CI_{\theta.AF} = [0.5846, 1.3940]$ with an interval length of 0.8094. Meanwhile, the confidence interval based on the Bayesian approach was $CI_{\theta.BS} = [0.5972, 1.3604]$ with an interval length of 0.7632. These results indicate that all of the confidence intervals contained the true ratio of the coefficients of variation. However, the Bayesian confidence interval provided the shortest length.

4.2. Example 2

To assess the PM2.5 level in Chiang Mai and Nan provinces, we used the data in Table 6 for the second analysis and summarized the statistics in Table 7. Using the Akaike Information Criterion values in Table 8, we found that the two PM2.5 samples came from log-normal populations.

The ratio of log-normal coefficients of variation for the Chiang Mai and Nan was $\hat{\theta} = 0.9654$. The confidence intervals for the ratio based on MOVER, modified MOVER, and approximate fiducial approaches were $CI_{\theta.MOVER} = [0.6355, 1.4667]$, $CI_{\theta.MMOVER} = [0.6153, 1.5031]$, and $CI_{\theta.AF} = [0.6171, 1.4988]$ with interval lengths of 0.8312, 0.8878, and 0.8817, respectively. Meanwhile, the confidence interval for the Bayesian approach was $CI_{\theta.BS} = [0.6274, 1.4457]$ with an interval length of 0.8183. The interval length of the Bayesian approach was shorter than the others, thus it more accurately estimated the coefficient of variation ratio for these two log-normal populations.

4.3. Example 3

The PM2.5 levels of Bangkok and Chiang Rai provinces in Table 9 were used to construct the confidence intervals for the ratio of coefficients of variation for comparing the dispersion of PM 2.5 with different levels. The statistics and the Akaike Information Criterion values were presented in Table 10 and Table 11, respectively. The result showed that the PM2.5 levels samples came from log-normal distributions.

The ratio of coefficients of variation of log-normal distributions for the Bangkok and Chiang Rai was $\hat{\theta} = 0.5300$. The confidence intervals for the ratio based on MOVER, modified MOVER, and approximate fiducial approaches were$CI_{\theta.MOVER}$ = [0.3519,0.7984], $CI_{\theta.MMOVER}$ = [0.3401,0.8130], and $CI_{\theta.AF}$ = [0.3411,0.8111] with interval lengths of 0.4465, 0.4729, and 0.4700, respectively. Moreover, the confidence interval for the Bayesian approach was $CI_{\theta.BS}$ = [0.3458,0.7901] with an interval length of 0.4443. The Bayesian confidence interval had the shortest interval length.

5. Discussion

Nam and Kwon (2017) proposed the MOVER approach for constructing the confidence intervals for the ratio of coefficients of variation of two log-normal distributions, while Hasan and Krishnamoorthy (2017) constructed them based on modified MOVER and approximate fiducial approaches and compared them with the MOVER approach. In this paper, we propose the Bayesian approach for the confidence interval estimation of the ratio of coefficients of variation of log-normal distributions.

6. Conclusions

Using the data examples from data set PM10 level and PM2.5 level in the northern Thailand, all approaches were illustrated with real data analysis. The performance of the Bayesian approach was compared to three existing approaches. The performances of the confidence intervals agreed with our simulation studies. Since the coverage probability of the Bayesian confidence interval was better than those of the others, and its average length was also shorter. Therefore, the Bayesian approach is recommended to construct the confidence intervals for the ratio of coefficients of variation of log-normal distributions when the dispersions of PM10 level and PM2.5 level are at the harmful level ($\geq 50\mu g/m^3$).

Acknowledgments

This research was funded by King Mongkut's University of Technology North Bangkok. Grant No. KMUTNB-63-DRIVE-13.

References

CÄSELLA, G., BERGER, R. L., (2002). *Statistical Inference*, California:Duxbury.

DONNER, A., ZOU, G. Y., (2002). Interval estimation for a difference between intraclass kappa statistics. *Biometrics*, 58, pp. 209–215.

ZOU, G.Y., DONNER, A., (2008). Construction of confidence limits about effect measures: a general approach. *Statistics in Medicine*, 27, pp. 1693–1702.

FAUPEL-BADGER, J.M., FUHRMAN, B.J., XU, X., FALK, R.T., KEEFER, L.K., VEENSTRA, T.D., HOOVER, R.N., ZIEGLER, R.G., (2010). Comparison of liquid chromatographytandem mass spectrometry, RIA, and ELISA methods for measurement of urinary estrogens. *Cancer Epidemiology Biomarkers & Prevention*, 19, pp. 292–300.

HANNIG, J., LIDONG, E., ABDEL-KARIM, A., IYER, H., (2006). Simultaneous fiducial generalized confidence intervals for ratios of means of lognormal distributions. *Austrian Journal of Statistics*, 35, pp. 261–269.

HARVEY, J., VAN DER MERWE, A. J., (2012). Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. *Journal of Statistical Planning and Inference*, 142, pp. 1294–1309.

HASAN, M. S., KRISHNAMOORTHY, K., (2017). Improved confidence intervals for the ratio of coefficients of variation of two lognormal distributions. *Journal of Statistical Theory and Applications*, 16, pp. 345–353.

KRISHNAMOORTHY, K., (2016). Modified normal-based approximation to the percentiles of linear combination of independent random variables with applications. *Communications in Statistics - Simulation and Computation*, 45, pp. 2428–2444.

KRISHNAMOORTHY, K., MATHEW, T., (2003). Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*, 115, pp. 103–121.

LACEY, L.F., KEENE, O. N., PRITCHARD, J. F., BYE, A., (1997). Common noncompartmental pharmacokinetic variables: are they normally or log-normally distributed? *Journal of Biopharmaceutical Statistics*, 7, pp. 171–178.

LIN, S.H., WANG, R. S., (2013). Modified method on the means for several log-normal distributions. *Journal of Applied Statistics*, 40, pp. 194–208.

MA, Z., CHEN, G., (2018). Bayesian methods for dealing with missing data problems. *Journal of the Korean Statistical Society*, 47, pp. 297-–313.

NAM, J.M., KWON, D., (2017). Inference on the ratio of two coefficients of variation of two lognormal distributions. *Communications in Statistics - Theory and Methods*, 46, pp. 8575–8587.

NIWITPONG, S.-A., (2013). Confidence intervals for coefficient of variation of lognormal distribution with restricted parameter space. *Applied Mathematical Sciences*, 7, pp. 3805–3810.

NG, C.K., (2014). Inference on the common coefficient of variation when populations are lognormal: A simulation-based approach. *Journal of Statistics: Advances in Theory and Applications*, 11, pp. 117–134.

RAO, K. A., D'CUNHA, J.G., (2016). Bayesian inference for median of the lognormal distribution. *Journal of Modern Applied Statistical Methods*, 15, pp. 526–535.

ROYSTON, P., (2001). The Lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Statistica Neerlandica*, 55, pp. 89–104.

SHARMA, M.A., SINGH, J.B., (2010). Use of Probability Distribution in Rainfall Analysis. *New York Science Journal*, 3, pp. 40–49.

TIAN, L., WU, J., (2007). Inferences on the common mean of several log-normal populations: The generalized variable approach. *Biometrical Journal*, 49, pp. 944–951.

THANGJAI, W., NIWITPONG, S.-A., (2019). Confidence intervals for the signal-to-noise ratio and difference of signal-to-noise ratios of log-normal distributions. *Stats*, 2, pp. 164–173.

THANGJAI, W., NIWITPONG, S.-A., NIWITPONG, S., (2016). Simultaneous fiducial generalized confidence intervals for all differences of coefficients of variation of log-normal distributions. *Lecture Notes in Artificial Intelligence*, 9978, pp. 552–561.

TSIM, Y. L., YIP, S. P., TSANG, K. S., LI, K. F., WONG, H. F., (1991). Haematogoy and Serology. In *Annual Report, Honk Kong Medical Technology Association Quality Assurance Programme*, pp. 25–40.

XING, Y. F., XU, Y. H., SHI, M. H., LIAN, Y. X., (2016). The impact of PM2.5 on the human respiratory system. *Journal of Thoracic Disease*, 8, E69–E74.

APPENDIX

Table 3: PM10 levels in Chiang Mai province and Nan province ($\mu g/m^3$)

Chiang Mai					Nan				
227	170	164	105	128	224	134	138	148	190
156	262	167	112	103	145	232	136	144	127
138	146	166	123	94	114	199	100	155	116
125	191	142	139	96	107	176	90	178	126
113	184	117	138	98	80	130	126	254	

Source: Pollution Control Department (http://aqmthai.com/aqi.php)

Table 4: Statistics of PM10 levels in Chiang Mai province and Nan province

Statistics	**Chiang Mai**	**Nan**
n	25	24
$\bar{y}$	144.1600	148.7083
s_Y	41.2580	44.9662
$\bar{x}$	4.9355	4.9603
s_X	0.2665	0.2931
$\hat{\tau}$	0.2656	0.2930

Table 5: The minimum Akaike Information Criterion values of PM10 level in Chiang Mai province and Nan province

Distribution	**Chiang Mai**	**Nan**
Normal	259.9186	253.7713
Log-Normal	**254.5765**	**250.2824**
Gamma	255.8663	250.9095
Exponential	299.5462	289.0954

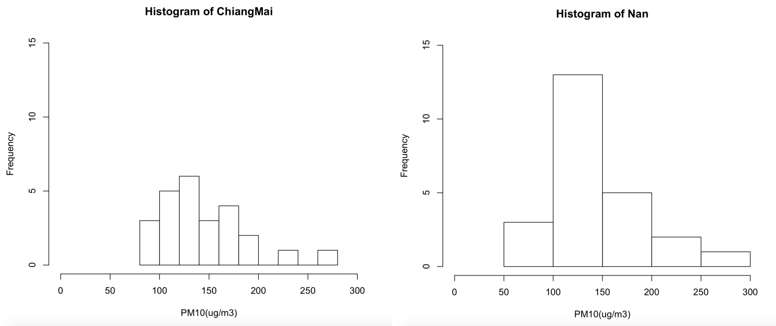

Figure 1: Histogram plots of PM10 level in Chiang Mai province and Nan province

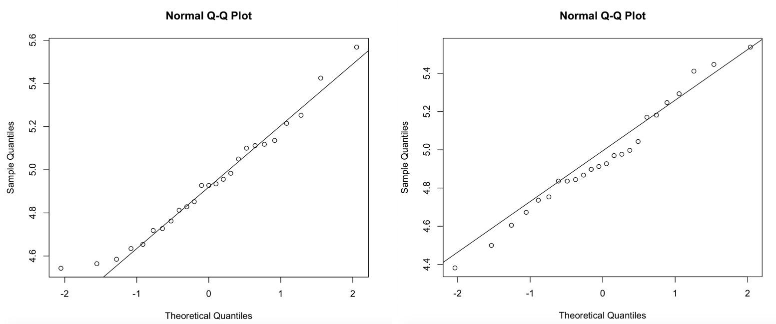

Figure 2: The normal QQ-plots of log-PM10 level in Chiang Mai province and Nan province

Table 6: PM2.5 levels in Chiang Mai province and Nan province ($\mu g/m^3$)

Chiang Mai					**Nan**				
189	129	124	69	92	192	104	111	115	154
118	213	126	72	68	118	199	107	108	100
100	109	125	83	64	88	167	73	119	90
92	147	105	99	66	86	146	61	136	89
82	145	79	102	62	55	105	96	209	

Source: Pollution Control Department (http://aqmthai.com/aqi.php)

Table 7: Statistics of PM2.5 levels in Chiang Mai province and Nan province

Statistics	**Chiang Mai**	**Nan**
n	25	24
$\bar{y}$	106.4000	117.8333
s_Y	38.1335	41.3718
$\bar{x}$	4.6120	4.7125
s_X	0.3324	0.3440
$\hat{\tau}$	0.3346	0.3465

Table 8: The minimum Akaike Information Criterion values of PM2.5 level in Chiang Mai province and Nan province

Distribution	Chiang Mai	Nan
Normal	255.9811	249.7724
Log-Normal	**249.4643**	**246.0677**
Gamma	250.9027	246.5411
Exponential	284.3603	277.9250

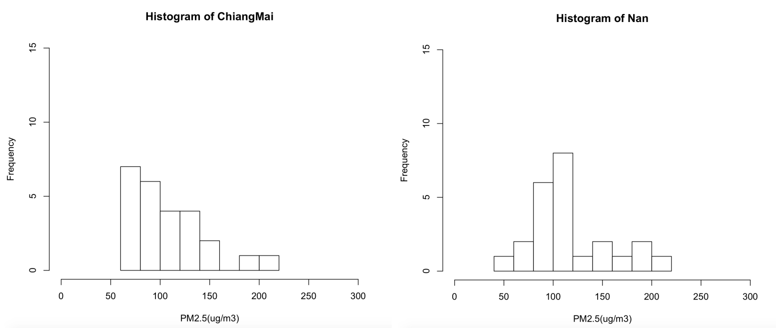

Figure 3: Histogram plots of PM2.5 level in Chiang Mai province and Nan province

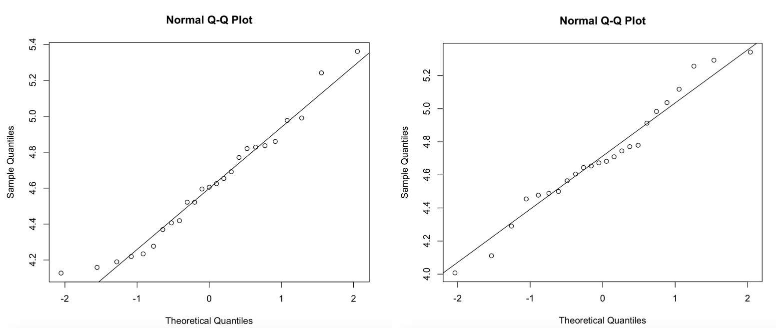

Figure 4: The normal QQ-plots of log-PM2.5 level in Chiang Mai province and Nan province

Table 9: PM2.5 levels in Bangkok province and Chiang Rai province ($\mu g/m^3$)

Bangkok					Chiang Rai				
30	19	18	25	19	184	89	109	63	104
22	19	21	15	14	147	228	77	72	85
22	23	15	16	14	79	254	77	79	74
20	19	22	16	15	77	140	83	82	113
20	23	17	18	13	86	132	82	104	162

Source: Pollution Control Department (http://aqmthai.com/aqi.php)

Table 10: Statistics of PM2.5 levels in Bangkok province and Chiang Rai province

Statistics	Bangkok	Chiang Rai
n	25	25
$\bar{y}$	19.0000	111.2800
s_Y	3.9791	49.8795
$\bar{x}$	2.9242	4.6361
s_X	0.2043	0.3762
$\hat{\tau}$	0.2022	0.3815

Table 11: The minimum Akaike Information Criterion values of PM2.5 level in Bangkok province and Chiang Rai province

Distribution	Bangkok	Chiang Rai
Normal	142.9793	269.4068
Log-Normal	**140.7307**	**256.8566**
Gamma	141.1844	260.2494
Exponential	198.2219	286.6025

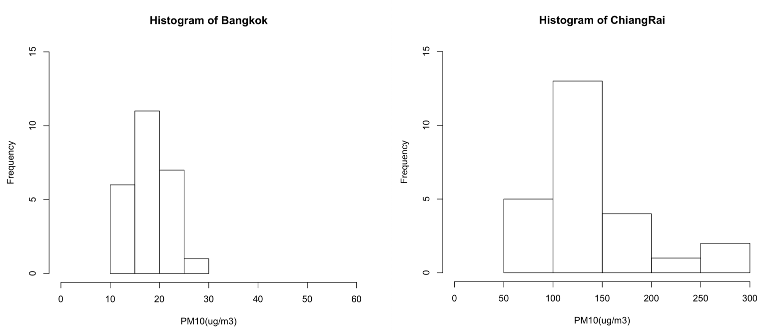

Figure 5: Histogram plots of PM2.5 level in Bangkok province and Chiang Rai province

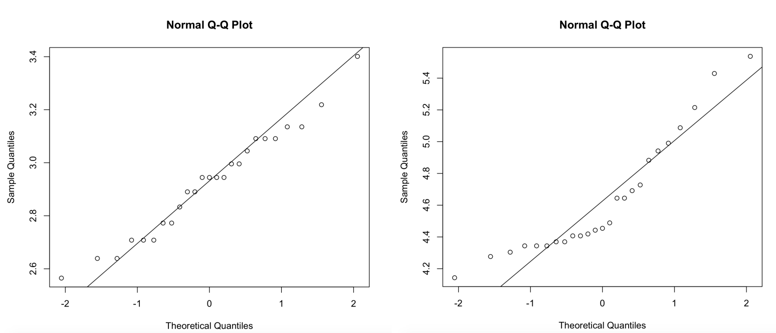

Figure 6: The normal QQ-plots of log-PM2.5 level in Bangkok province and Chiang Rai province