Kuwait J. Sci. **45 (2)** pp 22-29, 2018

Parametric and nonparametric bootstrap: an analysis of indoor air data from Kuwait

Sana BuHamra[1,*], Noriah Al-Kandari[2], Meshari Al-Harbi[3]

[1]*Dept. of Information Science, Kuwait University, Kuwait*

[2]*Dept. of Statistics and Operations Research, Kuwait University, Kuwait*

[3]*Dept. of Environmental Technology Management, Kuwait University, Kuwait*

**Corresponding author: sbuhamra@gmail.com*

Abstract

This paper discusses the performance of parametric and nonparametric bootstrap for confidence interval (CI) estimation applied to fine particulate matter ($PM_{2.5}$) data. Preceding the estimation process, several models were investigated to predict $PM_{2.5}$ concentrations from various tobacco smoking venues that resulted in a weighted logarithmic regression (WLS) model as a best fit. This model is then used as the base fit throughout the bootstrap estimation of the total number of burned cigarettes θ within an hour for a given a specific air quality level.

Keywords: Coverage proportion; cross-validation; nonparametric bootstrap; parametric bootstrap; $PM_{2.5}$.

1. Introduction

Modeling and predicting of fine particulate matter ($PM_{2.5}$) has recently gained increasing attention due to its sever health impacts (Pope III & Dockery, 2006) and the complexity of continuous measurement of its concentrations indoors. Statistical regression methods are the most common models available to predict outdoor and indoor air pollutant levels (Özbay, 2012). Cyrys *et al*. (2004) demonstrated the capability of multiple linear regression models to predict indoor $PM_{2.5}$, black smoke and particle number concentrations. Similar results were also reported by Elbayoumi *et al*. (2014). Tippayawong & Khuntong (2007) developed the mass-based balance model to predict $PM_{2.5}$ in Thailand schools. Balakrishnan *et al*. (2013) developed log-linear regression models that predict household $PM_{2.5}$ from solid cook fuel use. Evaluating model accuracy through *k*-fold cross validation revealed a reasonable degree of correlation (r=0.56) between modeled and measured values. Barakat *et al*. (2014) reported that block maxima and peak over threshold methods along with bootstrapping techniques resulted in goodness fit for outdoor emissions of SO_2, PM_{10} and O_3 in two Egyptian cities. Fuzzy logic technique was also used to develop environmental indicators for quantifying the environmental performance of industrial activities (Al-Shayji *et al*., 2008).

In 1995, smoking in all restaurants and other public places was banned in Kuwait; however, it was not implemented officially at that time. The Kuwait Ministry of Health reported (MOH, 2012) that the top three causes of death in the years 2008-2012 were cardiopulmonary diseases (62.3 per 100K inhabitants), neoplasms (22.4 per 100K inhabitants) and transport accidents (12.8 per 100K inhabitants). Thus, it is of great importance to deliver a quantifiable recommendation through some applicable actions to the Kuwait Environmental Protection Agency (KEPA) authorities in order to control smoking in public areas. Hence, the focus of the current work is to explore the best model for predicting $PM_{2.5}$ concentration Y based on the total number of burned cigarettes within a one-hour period, X. Subsequently, the model will be used to estimate $X = \theta$ at which the level of $PM_{2.5}$ becomes unhealthy based on 'US Idaho air quality health index for 1-hr $PM_{2.5}$ concentrations' developed by the US Idaho State Department of Environmental Quality (Table 1).

Table 1. Air quality health index rankings based on 1-hour $PM_{2.5}$ concentration developed by US Idaho State Department of Environmental Quality

1-Hour average $PM_{2.5}$ Conc. (μg/m^3)	Air Quality Index Category
0.0–40.0	Good
40.1–80.0	Moderate
80.1–175.0	Unhealthy for Sensitive Groups
175.1–300.0	Unhealthy
300.1–500.0	Very Unhealthy
500+	Hazardous

2. The $PM_{2.5}$ data

The TSI SidePak device was used to measure and record levels of $PM_{2.5}$ concentrations (μg/m^3) in selected eighteen cafes (Figure 1), distributed over all six governorates of Kuwait, that allow indoor smoking. Sampling was conducted during the periods of June-July and September-November of

2014, from 6-11 pm, where frequent visits to these venues were observed. Six criteria that allow indoor smoking were adopted for selecting the eighteen venues. These are closed-area smoking cafes that allow cigarette-smoking only (no cigar or water-pipes), ground floor, no major cooking activities, at least 50 percent occupied during the sampling period, and having well determined dimensions (area & volume). These criteria were chosen to control the effect of extraneous variables that might cause a confounded effect on the variable of interest (smoking). The data logging interval was set to one minute and the device was turned on and off inside each venue to prevent contamination with outside air. The air-monitoring device was positioned in a central location inside each venue. The duration of air sampling was 60 minutes for all the selected cafes. The number of people present and number of burning cigarettes were counted every six minutes. The final collected data consists of one response variable $PM_{2.5}$ concentrations (Y) and several explanatory variables such as: the total number of burned cigarettes within 1-hour testing period (X), venue volume (Volume), active smoking density defined as X divided by 100 m^3 (X_1) and number of people during the 1-hour testing period (N people). Most statistical analyses and simulations throughout this paper were conducted using software package R.

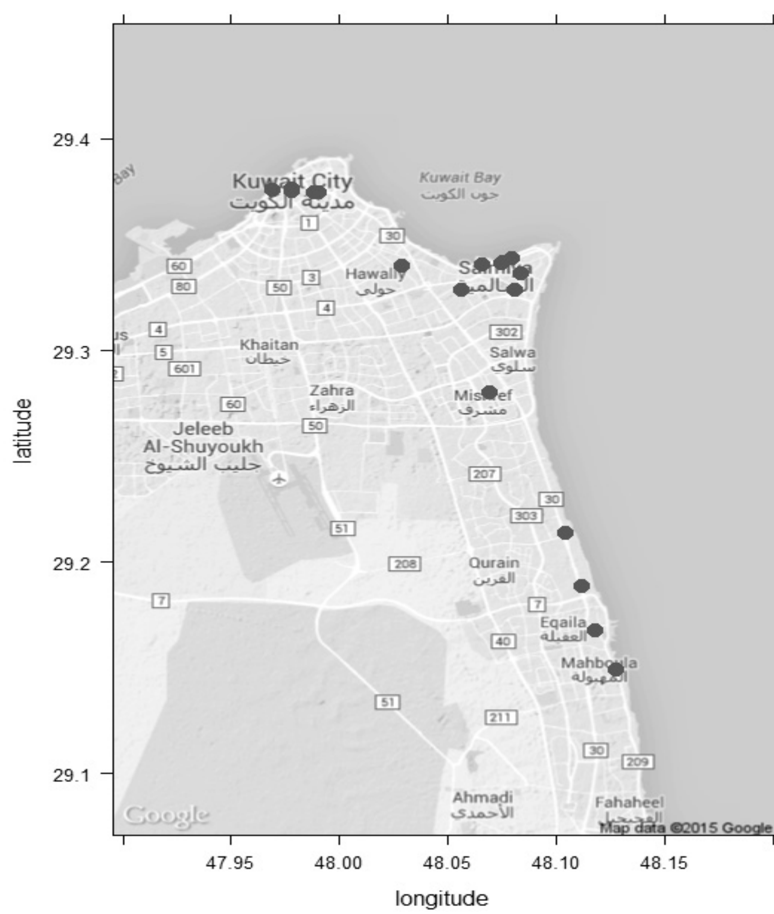

Fig. 1. Map of sampling sites

3. Model fitting

Preceding model building, an exploratory data analysis was performed. Table 2 provides descriptive statistics of the collected variables and correlations *r* between Y and the explanatory variables. The highest correlation was found between Y and X (r=0.812). Stepwise multiple regression analysis was employed and only X was found to be significant in predicting Y. The natural logarithmic transformation of Y (ln(Y)) was used to improve the model fit.

Several curves were investigated to fit ln(Y) on g(X), where the function g(X) was taken as linear, quadratic, and natural logarithmic in X. The weighted least squares (WLS) regression was also employed for fitting Y on X as well as ln(Y) on g(X), with weights taken based on the Volume. Table 3 compares different models in terms of Adjusted-R^2 and mean square error (MSE) criteria. The quadratic model (Model 3) showed lower performance and hence excluded from further analysis.

Residual analysis started with three main steps that were followed for the regression diagnostic in this study: 1) checking the assumptions of the model; 2) diagnosing outliers and influential points; and 3) cross-validation. Both

1 and Model 2 have comparable Adjusted-R^2. However, a close inspection of Figure 2 indicates a possible influential outlying point in the fit of Model 1. Diagnostic case statistics were performed on each of the 18 cafes by focusing on the three characteristics: discrepancy, leverage and influence. It can be seen from Figures 3(a)-3(b) that the standardized residuals are randomly scattered around the zero horizontal line. However, there is a point with standardized residual > 2 for Model 1 and therefore Model 1 is excluded from further attention.

Table 2. Descriptive statistics and correlations

Variables	Mean (Std. Deviation)	Correlations	
Y	229.61 (189.2)	*Y*	*ln*(*Y*)
X	18.89 (7.9)	0.81**	0.86**
Volume	397 (220)	0.50*	0.55*
X_1	7.03 (6.1)	0.44^{+}	0.32
N people	18.17 (12.5)	-0.20	-0.31

(**), (*), (+) correlation significant at level 0.01, 0.05 & 0.10, respectively.

For Model 2 in Table3, the relationship between ln(Y) and ln(X) satisfies linearity based on the lack of fit test (p-value=0.42). The normality assumption was validated using Shapiro-Wilks (p-value=0.8). The assumption of independence was verified using the Durbin-Watson test, which was found within the acceptable range from 1.5 to 2.5. The homoscedasticity assumption was validated using the residual plots (Figure 3(b)), and the Breusch-Pagan test (p-value=0.85).

Cross validation methods were used to estimate the prediction error and to identify the model with the lowest prediction error estimate(s). The data analysis and graphing (DAAG) package in R was employed in our program to conduct repeated cross validation. The 10-fold method as well as the 3-fold method (Harrell, 2013) were incorporated. The process was repeated 1000 times for each model in order to obtain a better prediction error estimate. It was found that both Model 2 and Model 5 have the minimum mean squared prediction error (MSPE) (10-fold CV=0.257) and lower than that of Model 4. However, in terms of Adjusted-R^2, Model 5 is better than Model 2. Consequently, based on both cross validation and residual analyses, Model 5 gives the best-fit for the $PM_{2.5}$ data under investigation.

4. Estimating θ

The main objective of this paper is to come up with a measurable indicator that helps maintain $PM_{2.5}$ concentrations within an acceptable air quality index (AQI) through controlling the number of burned cigarettes in indoor public venues. Hence, an estimation criterion is employed to obtain an estimator of $\theta = \theta(C, \pi)$ such that

$$\Pr(Y > C \mid X = \theta) \leq \pi,$$

where C is a given AQI limit as in Table 1 and π ∈{0.10,0.05,0.01} is given. Using Model 5, the solution of *ln(θ)* is given by

$$\hat{\theta} = \frac{ln(C) - b_0 - \hat{\sigma} z_\pi}{b_1}, \quad (1)$$

where z_π is the π^{th} percentile of the standard normal distribution. The coefficients b_0, b_1 and $\hat{\sigma}^2$ are the intercept, slope and the MSE of Model 5, respectively. Hence the estimate of θ is

$$\tilde{\theta} = \lfloor \exp(\hat{\theta}) \rfloor, \quad (2)$$

where $\lfloor . \rfloor$ is the floor function.

Table 3. Model summary and regression equation

Fit	Adjusted-R^2	MSE	Model Equation
Linear	0.727	0.243	Model 1: *ln*(*Y*)=3.116+0.103 *X*
Logarithmic	0.730	0.241	Model 2: *ln* (*Y*)=0.320+1.668 *ln* (*X*)
Quadratic	0.721	0.249	Model 3: *ln* (*Y*)=2.685+0.156 *X* -0.001 X^2
WLS 1	0.824	0.705*	Model 4: *ln* (*Y*)=0.106+3.025 *X*
WLS 2	0.822	0.713*	Model 5: *ln* (*Y*)=0.286+1.681 *ln* (*X*)

* weights are embedded in MSE

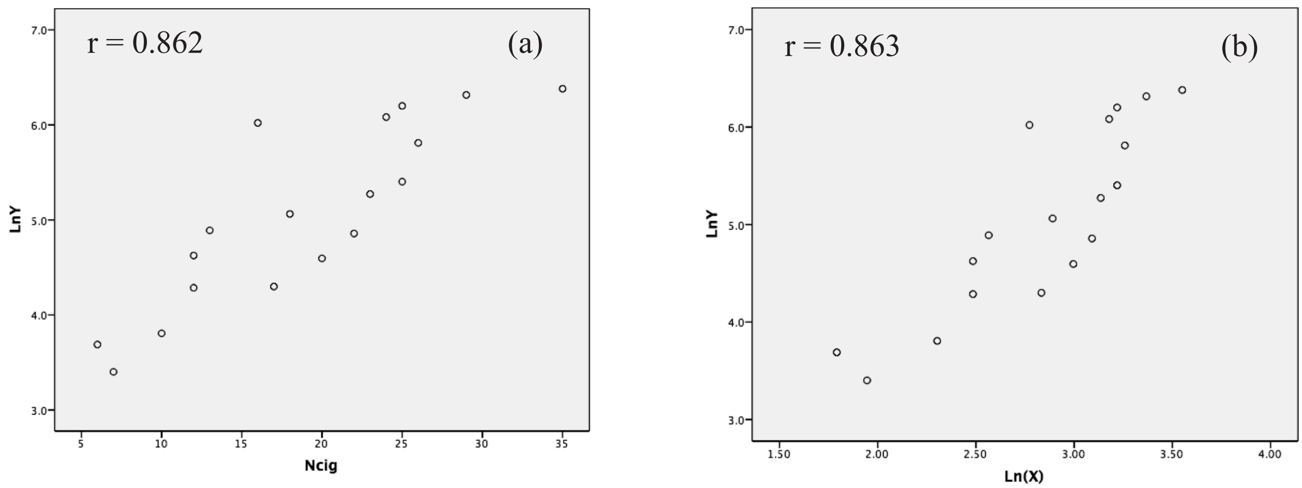

Fig. 2. (a) Plot of lnY versus X (b) Plot of lnY versus lnX.

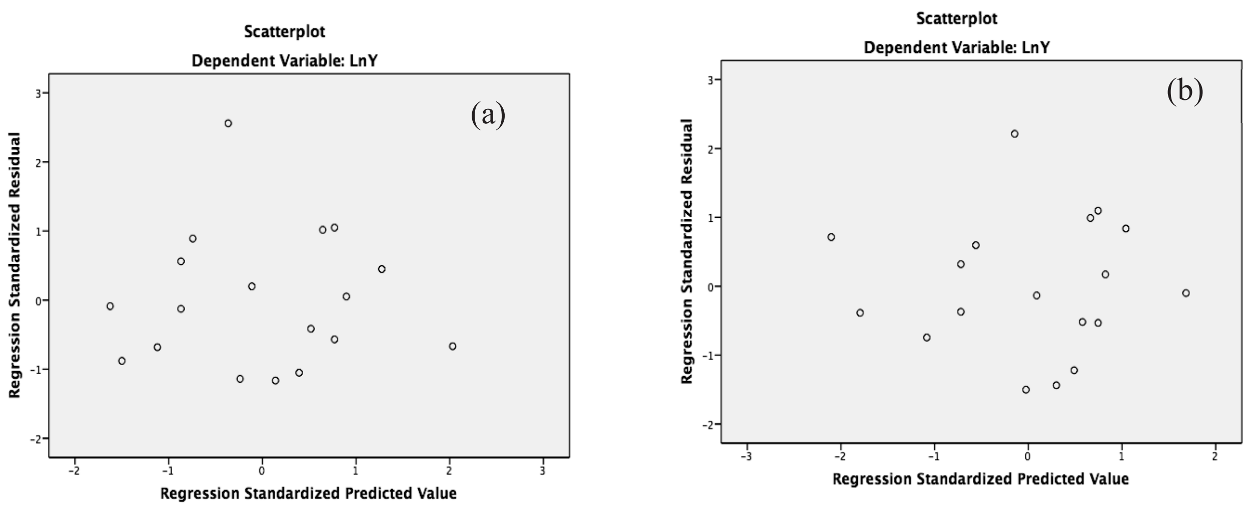

Fig. 3. (a) Residual plot Model 1 (b) Residual plot Model 2

Parametric and nonparametric bootstrap confidence interval estimation methods are investigated and compared (see for example, Davison & Hinkley, 1997; Chernick & LaBudde, 2014). The bootstrap estimation (Fox, 2002) of θ is computed based on Model 5 and the original data set. The number of bootstrap samples is m=1999 and consequently m estimators are generated in a vector $[\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(m)}]'$, where $\hat{\theta}^{(j)}$ is computed from the j^{th} sampled data of size n. The bootstrap estimator of $ln(\theta)$ and its standard error are given by

$$\theta_B = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}^{(j)}, \tag{3}$$

$$se(\theta_B) = \sqrt{\sum_{j=1}^{m}(\hat{\theta}^{(j)} - \theta_B)^2/(m-1)}. \tag{4}$$

4.1. Nonparametric bootstrap

The following bootstrap confidence intervals (CI) of $ln(\theta)$ will be used.

1. *Normal:*

$$(2\hat{\theta} - \theta_B) \mp z_{1-\frac{\alpha}{2}}\, se(\theta_B) \tag{5}$$

2. *Percentile* 1 (*P1*):

$$\left[\hat{\theta}_{(lower)} \quad , \ \hat{\theta}_{(upper)}\right] \tag{6}$$

3. Bias-corrected accelerated percentile (*BCa_1*):

$$\left[\hat{\theta}_{(adj\_lower)} \quad , \ \hat{\theta}_{(adj\_upper)}\right] \tag{7}$$

4. *Percentile* 2 (*P2*):

$$\left[2\hat{\theta} - \hat{\theta}_{(upper)} \quad , 2\hat{\theta} - \hat{\theta}_{(lower)}\right] \tag{8}$$

5. (*BCa_2*):

$$\left[2\hat{\theta} - \hat{\theta}_{(adj\_upper)} \quad , 2\hat{\theta} - \hat{\theta}_{(adj\_lower)}\right], \tag{9}$$

where $\hat{\theta}$ is as in Equation (1) and $\hat{\theta}_{(1)} < \hat{\theta}_{(2)} < \cdots < \hat{\theta}_{(m)}$ are the ordered values of $\hat{\theta}^{(j)}, j = 1, \ldots, m$. The desired estimator of the parameter of interest is then taken by applying the exponential function (e^x) on the limits of each confidence interval.

Table 4 lists the results of the nonparametric bootstrap estimation including the values of the bootstrap estimator θ_B, its standard error $se(\theta_B)$, the estimator $\tilde{\theta}$, and the five bootstrap confidence intervals from Equations (5) - (9) at 95% level. Other CI levels were also computed but not reported due to similarity of findings. The estimation performance in terms of minimum averaged interval length are ranked from best to worst as follows: $P2 \leq Normal < P1 < BCa\_1 < BCa\_2$. The proposed *P2* reveals that the total number of burning cigarettes within 1-hr should range between 5 and 9 cigarettes in order to maintain the concentration of $PM_{2.5}$ less than 80 µg/m^3. If the number of burned cigarettes is above 24, this implies hazardous AQI.

4.2. Parametric bootstrap estimation

The nonparametric bootstrap makes no assumptions about the underlying distribution and re-samples observations from the original data. The parametric bootstrap, nevertheless, generates the bootstrap samples from a given distribution as explained next. For *i*=1, 2,…, *m*=1999, we generate $\epsilon_j^* \sim N(0, MSE)$, *j*=1,…,18 and X^*'s are selected with replacement from the original *X* vector. The $Y_j^*{}'s$ are computed using the formula

$$ln(Y_j^*) = b_0 + b_1\, ln(X_j^*) + \epsilon_j^*,$$

where b_0, b_1 and MSE are computed from fitting Model 5 on the original data. Mimicking Equations (3) - (4), we denote the parametric bootstrap estimator and standard error, respectively, as θ_{Bp}, and $se(\theta_{Bp})$. When replacing θ_B, with θ_{Bp}, Equations (5) - (9) will provide the *Normal, P1, BCa_1, P2*, and *BCa*_2 parametric-bootstrap CIs.

Table 4. Nonparametric bootstrap estimation

C	π	θ_B	se(θ_B)	$\tilde{\theta}$	CI				
					Normal	***P2***	***P1***	***BCa_2***	***BCa_1***
80	.10	2.00	0.139	8	[6, 10]	[6, 10]	[6, 10]	[6, 11]	[5, 9]
	.05	**1.92**	**0.154**	**7**	**[5, 9]**	**[5, 9]**	**[5, 10]**	**[6, 10]**	**[5, 9]**
	.01	1.76	0.189	6	[4, 8]	[4, 8]	[4, 9]	[5, 13]	[2, 8]
500	.10	3.09	0.104	23	[17, 26]	[17, 26]	[19, 28]	[19, 28]	[17, 25]
	.05	**3.01**	**0.116**	**21**	**[15, 24]**	**[16, 24]**	**[17, 26]**	**[17, 34]**	**[12, 24]**
	.01	2.85	0.137	18	[13, 21]	[13, 21]	[14, 23]	[14, 26]	[11, 21]

Table 5 lists the results of the parametric bootstrap estimators and five bootstrap confidence intervals at 95% level. In general, the *P1* and Normal show minimum averaged interval length for *C*=80, whereas for *C*=500 the Normal and *P2* are the best. At error size π = 5%, the Normal CI shows that $\tilde{\theta}$ ranges between 5 and 11 cigarettes in order to maintain the concentration of $PM_{2.5}$ less than the moderate stage of AQI, while when $\tilde{\theta} > 25$ cigarettes indicating the hazardous stage. It is also noted that all five CIs have larger average length than those in the case of nonparametric estimation. Further comparisons between the two methods are investigated in the next section.

5. Simulation analysis

We generated N=1000 samples of size *n* from the model

$$ln(Y) = \beta_0 + \beta_1\, ln(X) + e,$$

where $e \sim N(0, \sigma^2)$, σ^2=0.22, $\beta_0 = 0.25$, $\beta_1 = 1.70$ and *X* was randomly selected with replacement from the original *X* vector.

Note that from Equation (2) the true value of θ is

$$\theta = \left\lfloor \exp\left(\frac{ln(C) - \beta_0 - \sigma z_\pi}{\beta_1}\right) \right\rfloor.$$

Table 5. Parametric bootstrap estimation

C	π	θ_{BP}	se(θ_{BP})	$\tilde{\theta}$	CI				
					Normal	***P2***	***P1***	***BCa_2***	***BCa_1***
80	.10	1.96	0.224	8	[5, 12]	[6, 12]	[5, 10]	[6, 16]	[4, 10]
	.05	1.87	**0.229**	**7**	**[5, 11]**	**[5, 12]**	**[4, 10]**	**[5, 16]**	**[3, 9]**
	.01	1.71	0.272	6	[4, 10]	[4, 11]	[3, 9]	[5, 18]	[2, 8]
500	.10	3.09	0.125	22	[17, 27]	[17, 27]	[18, 29]	[18, 30]	[16, 26]
	.05	**3.00**	**0.134**	**21**	**[15, 25]**	**[15, 25]**	**[16, 27]**	**[16, 31]**	**[13, 24]**
	.01	2.83	0.160	17	[12, 22]	[12, 23]	[13, 23]	[14, 40]	[7, 21]

For each of the 1000 data sets, we estimated θ and the nonparametric and parametric bootstrap CIs of Tables 4 and 5. These results were used to compute the average interval length and coverage proportion of each methods used in Tables 4 and 5. The results of this study are reported in Table 6. The results of Table 6 suggest that:

1. For nonparametric and parametric bootstrap when n=20 and C=80, the *BCa* _1, Normal and *P2* are more or less similar in terms of average length and coverage proportion with *BCa* _1 slightly better. For C=500 and n=20, both the Normal and *P2* perform better than the rest. For n=50, the *BCa* _1, Normal and *P2* perform better than the rest.

2. For n=20, the nonparametric bootstrap has, in general, shorten average length and slightly less coverage than the parametric bootstrap. For n=50, the parametric bootstrap performs better than the nonparametric bootstrap.

6. Conclusion

We considered the data on $PM_{2.5}$ concentrations in a sample of 18 cafés in Kuwait. The WLS model was used to fit the data. Nonparametric and parametric bootstrap methods were used to estimate the total number of burned cigarettes within one hour for a specific AQI level. The results suggest that to maintain the $PM_{2.5}$ concentrations within an acceptable or moderate air quality level, the total number of burning cigarettes should not exceed 9 cigarettes in an hour. A number above 24 burning cigarettes per 1-hr would most likely leads to a hazardous stage of air pollution.

7. Acknowledgments

Authors would like to thank Kuwait University for funding this research, Grant No. RQ02/13.Special thanks go to Prof. Partha Lahiri, University of Maryland, and Prof. Walid Bouhamra, Kuwait University, for their help during the early stage of the study. We also wish to thank Prof. Emad-Eldin Aly, Kuwait University for his constructive comments that substantially improved the presentation of the manuscript.

Table 6. Average CI length (coverage proportion) for nonparametric (NP) and parametric (P) bootstrap simulations

Bootstrap Method	n	C	π	*Normal*	*P2*	*P1*	*BCa_2*	*BCa_1*
NP	20	80	.10	4.19 (91.3)	4.07 (90.3)	3.80 (87.9)	6.09 (84.3)	4.01 (92.2)
			.05	4.27 (92.3)	4.20 (91.2)	3.90 (88.3	14.16 (84.4)	4.08 (93.0)
			.01	4.20 (91.9)	4.19 (90.4)	3.95 (86.5)	10.09 (82.8)	4.04 (93.7)
		500	.10	9.51(93.6)	8.39(92.8)	9.62 (85.6)	8.61(88.3)	8.22(90.2)
			.05	8.96(92.0)	8.10(91.1)	9.23 (82.4)	8.61(86.1)	8.01(89.1)
			.01	8.60 (91.6)	8.03 (92.3)	9.15 (81.4)	15.50(84.3)	8.20 (88.9)
	50	80	.10	2.34(92.3)	2.38(92.1)	2.35 (91.1)	2.62(87.4)	2.40(93.4)
			.05	2.39(92.0)	2.43(91.4)	2.40 (90.4)	2.72(87.5)	2.45(93.3)
			.01	2.44(92.1)	2.48(92.3)	2.45 (90.7)	2.85(87.6)	2.49(93.8)
		500	.10	5.29(93.7)	5.26(94.0)	5.48 (90.4)	5.40(91.4)	5.23(93.1)
			.05	5.18(95.0)	5.16(94.9)	5.38 (90.3)	5.38(91.1)	5.15(94.0)
			.01	5.03(93.1)	5.02(93.4)	5.23 (88.5)	5.43(88.4)	5.08(92.1)
P	20	80	.10	4.11(92.4)	4.32(91.0)	3.93 (92.3)	5.40(86.2)	4.11(95.3)
			.05	4.20(92.6)	4.43(91.3)	3.99 (91.8)	5.74(84.9)	4.17(95.2)
			.01	4.22(92.1)	4.48(90.2)	4.01 (91.4)	6.12(84.6)	4.14(95.3)
		500	.10	8.93(93.4)	8.87(94.6)	9.57 (88.7)	9.58(90.2)	8.91(91.9)
			.05	8.71(94.0)	8.72(94.6)	9.33 (88.5)	9.91(88.8)	8.83(93.6)
			.01	8.57(93.9)	8.67(94.2)	9.10 (88.3)	10.84(84.8)	8.89(93.6)
	50	80	.10	2.43(93.4)	2.47(92.1)	2.41 (93.0)	2.67(89.2)	2.45(94.9)
			.05	2.47(93.7)	2.52(93.9)	2.45 (93.1)	2.76(90.8)	2.50(93.4)
			.01	2.51(94.8)	2.57(93.2)	2.49 (93.3)	2.86(89.7)	2.53(94.2)
		500	.10	5.37(94.1)	5.37(94.7)	5.51 (92.1)	5.55 (92.5)	5.38(93.8)
			.05	5.28(94.7)	5.29(94.8)	5.42 (93.5)	5.57 (93.0)	5.33(94.2)
			.01	5.27(94.4)	5.30(94.5)	5.40 (91.3)	5.74 (89.4)	5.36(95.1)

References

Al-Shayji, K., Lababidi, H. M. S., Al-Rushoud, D. & Al-Adwani, H. A. (2008). Development of a fuzzy air quality performance indicator. Kuwait Journal of Science, **35**(2):101-126.

Barakat, H.M., Nigm, E.M. & Khaled, O.M. (2014). Statistical modeling of extremes under linear and power normalizations with applications to air pollution. Kuwait Journal of Science, **41**(1):1-19.

Balakrishnan, K., Ghosh, S., Ganguli, B., Sambandam, S., Bruce, N. ***et al.*** **(2013).** State and national household concentrations of PM2.5 from solid cookfuel use: Results from measurements and modeling in India for estimation of the global burden of disease. Environmental Health, **12**(1)77:1-14.

Chernick, R. & LaBudde, A. (2014). An introduction to bootstrap methods with applications to R. John Wiley & Sons.

Cyrys, J., Pitz, M., Bischof, W., Wichmann, E. & Heinrich, J. (2004). Relationship between indoor and outdoor levels of fine particle mass, particle number concentrations and black smoke under different ventilation conditions. Journal of Exposure Science and Environmental Epidemiology, **14**(4):275–283.

Davison, C. & Hinkley, V. (1997). Bootstrap methods and their application, volume 1. Cambridge University Press.

Elbayoumi, M., Ramli, A., Yusof, M., Yahaya, B., Al Mad-houn, W. ***et al.*** **(2014).** Multivariate methods for indoor pm 10 and pm 2.5 modelling in naturally ventilated schools buildings. Atmospheric Environment, **94**:11–21,

Fox, J. (2002). Bootstrapping regression models. URL http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-bootstrapping.pdf.

Harrell, E. (2013). Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer Science & Business Media.

Ministry of Health (MOH), State of Kuwait, Annual Report. (2012). Edition XLIX, Health, Kuwait.

Özbay, B. (2012). Modeling the effects of meteorological factors on SO2 and PM10 concentrations with statistical approaches. Clean–Soil, Air, Water, 4**0**(6):571–577.

Pope III, C. A. & Dockery, D.W. (2006). Health effects of fine particulate air pollution: lines that connect. Journal of the air & waste management association, **56**(6):709–742.

Tippayawong, N. & Khuntong, P. (2007). Model prediction of indoor particle concentrations in a public school classroom. Journal of the Chinese Institute of Engineers, **30**(6):1077–1083.

Submitted: 23/11/2015

Revised : 19/04/2017

Accepted : 20/04/2017

التقدير الإختزالي المعلمي واللامعلمي:
تحليل بيانات الهواء الداخلية في الكويت

سناء بوحمرا ، نوريه الكندري ، مشاري الحربي

قسم علوم المعلومات، قسم الإحصاء وبحوث العمليات، قسم إدارة التقنية البيئية

جامعة الكويت، الرمز البريدي13060،5969 الصفاة، الكويت

الخلاصة

تناقش هذه الورقة أداء أساليب التقدير الإختزالي (البوتستراب) المعلمي واللامعلمي لتقدير فترات الثقة مطبقة على بيانات تركيزات الجسيمات الدقيقة$PM_{2.5}$. قبل عملية التقدير، تم دراسة العديد من النماذج لإيجاد أفضل نموذج للتنبؤ بتركيز $PM_{2.5}$ في عدة أماكن يسمح فيها بتدخين التبغ، مما أسفرت عن ترجيح نموذج انحدار لوغاريتمي (WLS) على أنه الأكثر ملائمة. وتم إعتماد هذا النموذج خلال تقديرات بوتستراب لإجمالي عدد السجائر المحروقة في غضون ساعة عند مستوى معين لجودة الهواء.

الكلمات المفتاحية: الجسيمات الدقيقة $PM_{2.5}$ ؛ جودة الهواء؛ البوتسراب اللامعلمي؛ البوتسراب المعلمي؛ إحتمال التغطية؛ التحقق من المصادقة.