Statistical Study of PM2.5 Measured at the Kamloops Federal Building for Years 2011, 2012 and 2013

prepared by
Dr. Peter Tsigaris, Professor of Economics, School of Business and Economics,
Thompson Rivers University
and
Dr. Robert S. Schemenauer, Ph.D. (Atmospheric Physics, Retired)
for
Kamloops Physicians for a Healthy Environment Society
9 July 2014

Abstract

This study was conducted to examine the annual mean $PM_{2.5}$ values, as determined from the daily average values, at the Federal Building in Kamloops for the period 2011-2013. The work discussed here examines the air quality data and statistical analyses used by KPHES to report on $PM_{2.5}$ values for years 2011, 2012 and 2013 as stated in the society's June brochure. The evidence provided in this study indicates, with high confidence, that the 2013 annual mean $PM_{2.5}$ value exceeds the B.C. guideline level of 8 μg/m^3. Our analyses also show that the annual mean (and median) value for 2013 is greater, at the 95% confidence level, than the mean (and median) values for both 2012 and 2011.

Introduction and Overview of the Data

The discussion below refers to $PM_{2.5}$, which is airborne particulate matter with aerodynamic diameters equal to or less than 2.5 micrometers. The daily average values (daily values) from January 1st 2011 to December31st 2013 were downloaded from the B.C. air quality website in order to examine air quality in Kamloops. All data were obtained from the modern beta attenuation monitoring (BAM) instrument at the Federal Building in downtown Kamloops. Daily values were produced by averaging the 24 hourly measurements for each day. We used the daily average values in our calculations; daily average values are also used in health advisory notices by Environment Canada. Computing the annual mean from daily average values or from hourly data will give identical results for the annual average figure reported. The current 24 hour average guideline in B.C. states that the daily average value should be no higher than 25 μg/m^3. This daily average value is different from the yearly average value, which should be no higher than 8 μg/m^3 .

Table 1 below provides some descriptive statistics on the daily average values of $PM_{2.5}$ for the years 2011, 2012, and 2013. The first column shows the year, the second column indicates the number of valid daily observations (N) and the third column displays the number of missing values (N*). For these days no data are available from the BC Ministry of the Environment (BC MOE). The column labelled “mean” shows the average annual $PM_{2.5}$ values which are the ones reported in the KPHES brochure. The 2012 and 2013 average values are the same as those reported in in the B.C. Lung Association 2013 and 2014 reports. This agency did not report a Kamloops value for 2011.

Table 1: Summary of Statistics for Years 2011 to 2013

Year	N	N*	Mean	St Dev	SE Mean	Q1	Median	Q3	Q98
2011	363	2	7.6	4.0	0.21	4.9	6.9	9.3	18.5
2012	357	9	8.0	4.2	0.22	5.0	6.8	10.2	19.3
2013	355	10	8.9	4.6	0.24	5.6	7.8	11.2	21.3

Notes: Descriptive statistics for the daily average values. See text for an explanation of the terms.

Next in the table is the standard deviation (St Dev) of the daily average value over the year. The standard deviation describes how spread-out the daily values are relative to the mean. The standard error of the sample mean is shown next in the table (SE Mean). This indicates the amount of variation in the sample average under repeated sampling, i.e. how far the calculated (measured) mean could be from the population (actual) mean. It is equal to the standard deviation of the daily values (St Dev) divided by the square root of the number of observations in the sample. For example, in 2013 the standard deviation of the daily average values is 4.6 $\mu g/m^3$ and the standard error of the sample mean is 0.24 $\mu g/m^3$ (i.e., $4.6/\sqrt{355}$)). There are 355 observations in 2013 that are used to compute the mean and the standard error of the sample mean. The standard error of the sample mean is a low value in all three years, which is “good” statistically speaking. It is low because there are a large number of data points used. A low standard error of the sample mean indicates that there is not a lot of variation in the sample average under repeated sampling. If another sample was taken at a station close to the current station, for example, the sample mean would be very close to the current value.

The median is the value with half the observations less than this value and the other half more than this value, i.e. it is the middle value. For example, in 2013, 50 percent of the observations were below 7.8 and 50 percent were above 7.8 $\mu g/m^3$. The median is often used to make statistical inferences when the distribution is skewed. In a skewed distribution, the data are not distributed “normally” about the mean, i.e. the data do not have the classic “bell-shaped” curve (see Figure 1). The median is not impacted much by extreme daily values, e.g. forest fires, while the sample mean is affected much more by extreme events. In this report we also examine the median because the distribution of daily values is slightly skewed. Q1 is known as the first quartile, with 75 percent of the data being greater than the value reported, and Q3 is the third quartile, where 25 percent of the data are greater than this value. In 2013 about 25% of the 355 daily values (about 3 months of the year) exceeded 11.2 µg/m^3. Q98 is the 98th percentile which indicates what is happening at the tail end of the distribution.

The mean values and the median values are increasing over the three years. The Q1, Q3 and Q98 values are increasing as well. This is a first indication that the annual distribution of the daily average $PM_{2.5}$ is shifting towards higher values. The 2013 distribution of the daily average $PM_{2.5}$ values is shown below in Figure 1.

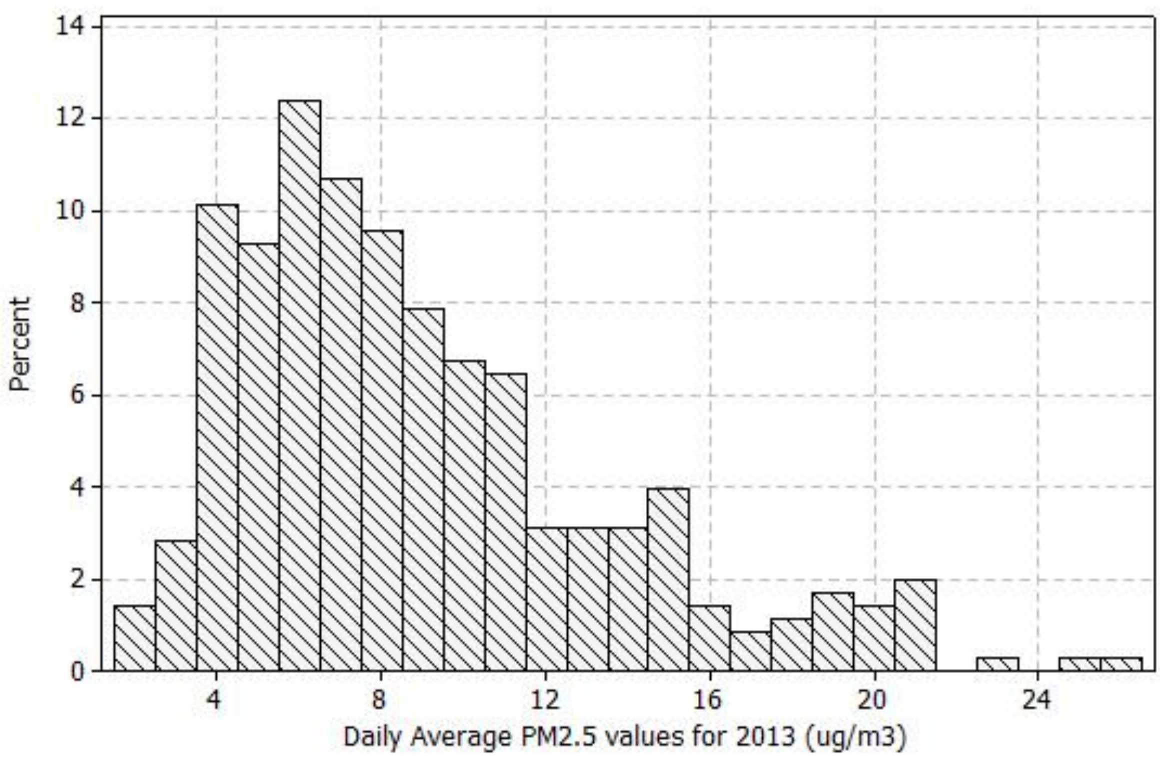

Figure 1: Histogram of the daily average PM2.5 values for the Kamloops Federal Building in 2013. The data were obtained from the British Columbia Ministry of the Environment.

There are not many days with daily average values equal to or less than 6 µg/m^3. On the opposite end, on January 15th 2013 the daily average value exceeded the B.C. guideline advisory value of 25 µg/m^3.

As Figure 1 shows, the data distribution is skewed lacking a perfect normal distribution. However, confidence intervals can be used with distributions that are not normal. The characteristics of the distribution are shown in Table 2, with 95% confidence intervals.

A 95 % confidence interval is constructed for the purpose of adding accuracy to the sample mean as an estimate of a true unknown population mean. The intervals show a low and a high bound for estimates of the unknown population mean. The 95 % confidence tells us the percentage of times that the confidence interval actually contains the population mean, under the assumption that the sampling is repeated a large number of times. It shows that 95 percent of the time (95 out of 100 times) the observed interval contains the true population mean but 5 out of 100 times the observed interval may not contain the true population value.

Table 2: The 95% confidence interval for the mean values using the t-distribution

Year	N	Mean	St Dev	SE Mean	95% confidence interval
2011	363	7.64	3.98	0.209	(7.23,8.05)
2012	357	7.96	4.15	0.219	(7.53,8.39)
2013	355	8.87	4.60	0.244	(8.39,9.35)

Notes: Description of the $PM_{2.5}$ sample distributions for years 2011, 2012 and 2013.

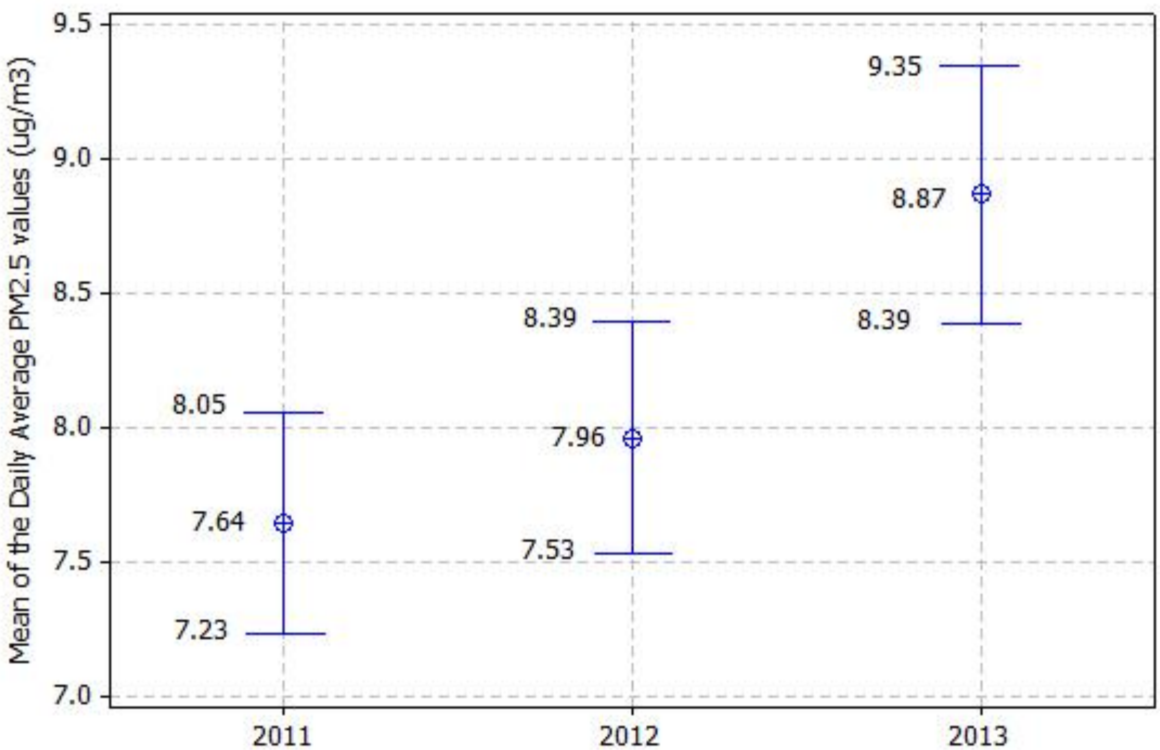

Figure 2: The 95 % confidence intervals for each of the three years (see text for more detail).

For example, a 95 percent confidence interval about the unknown population mean value of the $PM_{2.5}$ daily value in 2013 is constructed as follows. Taking the sample mean value of 8.87 $\mu g/m^3$ and adding two times the standard error of the sample mean yields the upper end value of the confidence band, while subtracting this same amount will give the lower end of the confidence band. Thus we have a very high confidence (95% confidence) that the annual average value is within the range 8.39 to 9.35 $\mu g/m^3$. This means we have strong confidence that the annual mean daily (average) value is above the BC Governments guideline value of 8 μg/m^3. Figure 2 shows the two standard error wide interval bands, above and below the mean values, for each of the three years.

What we learn from Figure 2 is that we can say, with 95% confidence, that the mean value of daily average $PM_{2.5}$ in 2013 is greater than the mean value for 2012 and greater than the mean value for 2011. We cannot say with the same confidence that the 2012 mean is greater than the 2011 mean value. This is discussed in much more detail below.

In Table 3 we see the frequency of occurrence of daily average values of $PM_{2.5}$ for the three years. The occurrences of different levels in 2012 and 2011 seem to be quite similar. $PM_{2.5}$ levels in 2013 year stand out as the frequency is higher for all $PM_{2.5}$ categories. In other words, there was a higher frequency of bad days and a lower frequency of good days in 2013 relative to 2012 and to 2011.

Table 3: Number of Times Different PM2.5 Levels Were Exceeded in Each Year

Year	Days	$PM_{2.5} \geq 25$	$PM_{2.5} \geq 20$	$PM_{2.5} \geq 15$	$PM_{2.5} \geq 10$	$PM_{2.5} \geq 8$	$PM_{2.5} \geq 6$
2011	363	0	4	23	77	138	225
2012	357	1	4	22	91	144	215
2013	355	1	12	37	116	173	252

Notes: Frequency of occurrence table for $PM_{2.5}$ for the three years. The number of times values above the various levels were measured in each year is given.

We use the sample mean and the standard error of the mean to make inferences about the unknown population mean of the daily $PM_{2.5}$ values and also to compare means across years. This is done in the next section.

Detailed Statistical Comparison of the Years 2011, 2012 and 2013

How different is 2013 from 2012? From 2011? Statistical tests can be used to determine if any differences are truly significant. Parametric tests assume that all data are distributed "normally" about the mean. There are other tests (non-parametric tests) that can be used if the data are not normally distributed. To be prudent, we have used both types of tests in our analyses. Table 4 compares the sample means for the different years using parametric t-tests. The first test compares the annual average mean $PM_{2.5}$ in 2013 with that observed in 2012. The estimated difference is 0.91 μg/m^3 and the 95 percent confidence range extends from 0.262 to as high as 1.55 ug/m^3 μg/m^3. The test statistics reveal that the difference between the two values, for years 2013 and 2012, is highly statistically significant. We can reject the null hypothesis that there is no difference between the two years with a 0.006 probability of making a decision error. We can conclude that 2013 had a statistically significant higher mean $PM_{2.5}$ daily average value than did 2012.

A similar conclusion can be stated for a comparison between 2013 and 2011. The year 2013 had a statistically significant higher mean $PM_{2.5}$ daily average value than the year 2011.

When we compare 2012 and 2011 we cannot conclude that there is a statistically significant higher mean daily average value in 2012 relative to 2011. There is a 29% probability of observing a difference at least this large by chance.

Table 4: Comparisons of two population means assuming unequal variances using t-tests.

```
Two-sample T for PM2.5 2013 vs PM2.5 2012

              N  Mean  StDev  SE Mean
PM2.5 2013  355  8.87   4.60     0.24
PM2.5 2012  357  7.96   4.15     0.22

Difference = mu (PM2.5 2013) - mu (PM2.5 2012)
Estimate for difference:  0.909
95% CI for difference:  (0.264, 1.553)
T-Test of difference = 0 (vs not =): T-Value = 2.77  P-Value = 0.006  DF = 701

Two-sample T for PM2.5 2013 vs PM2.5 2011

              N  Mean  StDev  SE Mean
PM2.5 2013  355  8.87   4.60     0.24
PM2.5 2011  363  7.64   3.98     0.21

Difference = mu (PM2.5 2013) - mu (PM2.5 2011)
Estimate for difference:  1.227
95% CI for difference:  (0.596, 1.859)
T-Test of difference = 0 (vs not =): T-Value = 3.82 P-Value = 0.000  DF = 696

Two-sample T for PM2.5 2012 vs PM25 2011

              N  Mean  StDev  SE Mean
PM2.5 2012  357  7.96   4.15     0.22
PM2.5 2011  363  7.64   3.98     0.21

Difference = mu (PM2.5_2012) - mu (PM25_2011)
Estimate for difference:  0.319
95% CI for difference:  (-0.276, 0.914)
T-Test of difference = 0 (vs not =): T-Value = 1.05  P-Value = 0.293  DF = 715
```

A paired-difference t-test was also conducted and the results remain intact. The mean for the year 2013 was significantly higher than those for the years 2012 and 2011. As for 2012 and 2011, no significant difference in the means for the two years was detected.

We also conducted the Mann-Whitney non-parametric test to examine the equality between two population medians (Table 5) and we reached the same conclusions. The year 2013 had a significantly different population median relative to 2012 and to 2011. The 2012 population median and that for 2011 were undistinguishable. Thus the hypothesis of no change in population medians cannot be rejected for 2011 and 2012.

Table 5: Mann-Whitney non-parametric test comparing two population medians

```
Mann-Whitney Test and CI: PM2.5 2013, PM2.5 2012

              N  Median
PM2.5 2013  355   7.800
PM2.5 2012  357   6.800

Point estimate for ETA1-ETA2 is 0.700
95.0 Percent CI for ETA1-ETA2 is (0.200,1.300)
W = 133944.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0071
The test is significant at 0.0071 (adjusted for ties)

Mann-Whitney Test and CI: PM2.5 2013, PM2.5 2011

              N  Median
PM2.5 2013  355   7.800
PM2.5 2011  363   6.900

Point estimate for ETA1-ETA2 is 1.000
95.0 Percent CI for ETA1-ETA2 is (0.400,1.500)
W = 137441.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0004
The test is significant at 0.0004 (adjusted for ties)

Mann-Whitney Test and CI: PM2.5 2012, PM2.5 2011

              N  Median
PM2.5 2012  357  6.8000
PM2.5 2011  363  6.9000

Point estimate for ETA1-ETA2 is 0.2000
95.0 Percent CI for ETA1-ETA2 is (-0.2997,0.7001)
W = 131199.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.3702
The test is significant at 0.3702 (adjusted for ties)
```

Conclusions

The evidence provided indicates, with high confidence, that the 2013 annual mean $PM_{2.5}$ value exceeds the B.C. guideline level of 8 $\mu g/m^3$. We found compelling evidence that the year 2013 had significantly higher daily average $PM_{2.5}$ population mean and median values than the mean and median values for each of the years 2012 and 2011. This rigorous study confirms the results presented in the June 2014 KPHES brochure.

We chose a full year as the time-frame for analyzing measurements because this is the standard in reporting $PM_{2.5}$. We do not speculate on the cause of this change. Nonetheless, something significant must have happened to impact 2013 that was not present in 2012 or in 2011. Wood stove usage is not likely to be the source because just as many, if not more, wood stoves were probably being used in 2012 and in 2011. The same applies to cars idling and other residential activities. In general, households are becoming more aware of the health problems of air pollution over the last decade and many are taking measures to reduce their emissions. Some industry may also have been reducing $PM_{2.5}$ since 2009. The year 2013 was also relatively low in terms of forest fire activity. There is no documented reason to believe more forest fire particulates were present in the city in 2013 than in other years but this idea will be examined further. Such a future study will specifically address the matter of Siberian forest fires in 2012 and whether they may or may not have had an effect on the 2012 data for $PM_{2.5}$. Examining the different seasons across the years can also provide new insights. We will continue to examine the data base for $PM_{2.5}$ in Kamloops and present further information on this subject in another report.

Acknowledgement

We would like to acknowledge three reviewers at Thompson Rivers University who examined the statistical data presented in this report. In particular, one reviewer spent considerable time assisting with the presentation of the data and we are especially grateful for the time and the insights.

Source for the Data

The data are all from the downtown Federal Building in Kamloops for the period from 1 January 2011 to 31 December 2013. The data were collected with one type of instrument at one location. The measurements were made by the British Columbia Ministry of the Environment, who also did the quality assurance and quality control of the data. The data are available via dropbox at: https://www.dropbox.com/s/aubpu0fnmcg8jzr/data_pm2.5_2011_2013.xlsx

General References (Not Cited in the Text)

B.C. Air Quality New Ambient Air Quality Criteria for PM2.5, 2014, accessed at: http://www.bcairquality.ca/regulatory/pm25-objective.html

Population data, British Columbia, "Is there a "safe" level of air pollution in British Columbia," page last revised: November 5, 2013, https://www.popdata.bc.ca/ria/casestudy/Lillooet_air_study

Engel-Cox, Jill, Nguyen Thi Kim Oanh, Aaron van Donkelaar, Randall V. Martin, and Erica Zell. "Toward the next generation of air quality monitoring: Particulate Matter." *Atmospheric Environment* 80 (2013): 584-590.

Lippmann, Morton. "Particulate matter (PM) air pollution and health: regulatory and policy implications." *Air Quality, Atmosphere & Health* 5, no. 2 (2012): 237-241.

Ministry of Healthy Living and Sport, "Guidance on application of provincial air quality criteria for PM2.5" June 2009, accessed fromhttp://www.bcairquality.ca/reports/pdfs/pm25-implement-guide.pdf

The Lung Association of British Columbia, Reports can be accessed at: http://www.bc.lung.ca/airquality/stateoftheair-report.html