Statistical Analysis for Equipment Usage in Electronic Classrooms using SAS Macro

Yong Xu[1] and Joshua Williams[2]

[1]Ferris State University, Marketing Department
Big Rapid, MI 49307, USA
[2]Radford University, Box 7004
801 E. Main Street, Radford, VA 24142

ABSTRACT: With modern industrialization and high technology, people tend to consume more and more energy. With world population increase very fast and living standard increasing rapidly, the challenge for energy saving is more and more demanding. We plan to start with a statistical analysis to one campus electronic devices usage to understand the pattern of the energy cost in classroom. We developed a SAS macro for people to use for similar research purpose. We wish with better understand of how people consume energy to eventually save significant amount of energy and have a better future. The technology and multimedia devices become more and more present in classroom environments from Elementary Schools to Universities. Understanding which devices and how often they are used is needed. We are using data collected from Crestron Room View machines in sixty eight multimedia classrooms to develop time series models. Having the statistical models of classroom technology/multimedia will aid IT Professionals, Administrators, and Teachers by allowing for more efficient classroom design and device selections. It provides one direction to help people to conduct big data or machine learning analysis for energy saving, usage effective and environmental protection purpose.

Key Words: ARIMA, Time Series, Energy Saving, SAS, SAS Macro, Big data, Machine Learning

1 INTRODUCTION

With modern industrialization and high technology, people tend to consume more and more energy. With world population increase very fast and living standard increasing rapidly, the challenge for energy saving is more and more demanding. Even the population regenerate rate is much lower in developed countries. With high population regenerate rate in a lot of developing counties, the overall world population regeneration rate is still increasing. Even the rate is low but considering the fact with modern medical and science and technologies improvement, the world average life expectancy is increasing. In other world, nowadays people live longer than their ancestors. Therefore, the world population is still growing and in a short time period (like 10 or 20 years), this trend will become more and more obvious. On one side, the number of people who needs energy is increasing. On the other hand, with most countries continuously improve their economy and people's living standards improve, the individual people's energy consumption is also increasing. With the two factors combined, the world energy demand is increasing rapidly in the last decade. More energy consumptions generate the need for more energy generation. With the limits of energy generation abilities, we have the conflict of energy demand and energy supply. It create the necessity of energy saving and more efficient use of energy. At the same time, with the usage of energy increasing, now we as human being create sever problem in environmental challenges such as global warming, global water and air quality, pollutions from trash in land, sea and air, etc. To enhance the efficiency of energy usage and lower the cost of energy usage as well as protecting the environment with less tree cost and less pollution, we want to have methods to better understand the energy usage pattern in order to save energy, reduce cost and protect our environment.

In this study, we will introduce a method that we used to understand equipment usage in a university class room. Nowadays, people have a lot of data available to them thanks to the modern data gathering and censoring technologies. With internet of things become available, we can anticipate more and more data can be available. However, one of the key difficulties for people to do research on these data is clean the data in a way such that people can perform follow-up research. We will introduce a SAS macro to help other people to be able to conduct similar research to understand more energy usage pattern. Once people can understand better then we can always optimize the usage so that we can save energy and reduce the cost. Many researches had been done in the usage understanding of classroom technology and machine leering of data.

2 LITERATURE REVIEW

Many studies on In Classroom Technology (ICT) focus primarily on how equipment is used by instructors or the instructor's perceptions towards ICT. Usluel, Aşkar & Baş (2008) looked at faculty access to In Classroom Technology (ICT) through a survey in Turkey. The survey asked what purpose the ICT was used and how long instructors had been using ICT. Usluel, Aşkar & Baş's survey found that 56% of the faculty listed the projector as the most common item available to them. Gülbahar (2008), also in Turkey conducted similar ICT usage in the classroom research. This study found that 58% of the faculty surveyed used computers for presentations. Brill and Galloway (2007) conducted a research of faculty

at a research university in the United States. They investigated the usage of classroom based teaching technologies and they summarized the attitudes of college level instructors toward teaching technologies. They found overhead projectors and VCR's to be the two devices with the highest usage among faculty in that university.

In energy usage of equipment studies, questions such as which kind of equipment is used and how often will be people's primary focus. Webber, et al (2005) and Roberson, et al (2004) studied the electrical usage of office equipment after business hours during the weekdays and weekends at 12 buildings (including office buildings, schools and hospitals). These researches had interest to observe which devices were being left on, which were in standby mode and which were turned off. Webber, et al. (2005) explained the uncertainties that come from estimating energy usage, user's behavior and power management rates. Compared with previous researches, the improvement of this type of research is that they have more accurate data. The problems with this research or rather the limitation of this research which prevent it from widely accepted by many other places are simply the costs of the research. In other words, it cost too much money. For example, if you have to spend more money to do this type of investigations which lead to the potential benefit of saving money (less than the money spend to hire professional personals to visit and track all the data) on equipment and power cost then you probably will pass it by.

These studies and others like them often ask what kinds of equipment are used and how often as the secondary question. One problem with these researches is people use objective data rather than subjective data. For example, in most researches, researchers will ask participant give answers to the question such as rank the usage of the following equipment or which equipment is used more. People can only relay on their memory or personal intuition to give answers. Sometimes, people are not able to answer questions very accurate because of forgetting or unsure.

To get more accurate and detailed data on what equipment is used and how often, surveys such as these would be not good enough. This is due to the subjective measurements and the dependency on accurate responses in sufficient numbers to generate a model.

Later scientists do notice the problem and they try to use equipment feedback data to conduct their research. For example, Chow and Wu (2019) have interested to evaluate the effectiveness of Outdoor fitness equipment (OFE) to help policy maker to develop Future Park or open space initiatives. OFE areas have become a necessary infrastructure in public open spaces as a means to improve public health through increased physical activity for people. They use data which directly gathered from the equipment rather than just survey questions. Choi, et al (2018) conducted the research on smart cities. They studied the intelligent power equipment management systems to form sustainable and effective energy systems by understanding energy efficiency. They also use the data which are collected directly from machines rather than from survey.

With the data from RoomView machines, we conduct accurate and efficient analysis to understand which ICT was used and how often these ICT are used so that we can help policy makers to make smart decisions. Autoregressive Integrated Moving Average (ARIMA) modeling has been widely with many researchers in the last several decades. Bowerman, et al (2005) studied forecasting and time series problems. Shenston and Hyndman (2005) utilize stochastic models to estimate intermittent demands. Shih and Tsokos (2008) applied ARIMA models to analyze Carbon dioxide concentration to better understand global warming issues. Hyndman and Khandkar (2008) developed a forecasting package using R language. Shih and Tsokos (2009) forecasted the CO_2 amount in the atmosphere through ARIMA. Abdullah (2012) applied ARIMA modeling for prediction of gold selling prices. Vasileiadou and Vliegenthart (2013) studied dynamic social processes using an ARIMA model. Williams and Xu (2014) use Time Series Analysis to conduct ARIMA forecast for Device Usage in Electronic Classrooms. Xu and Choi (2018) conduct a research use machine learning to understand breast cancer problems. In this paper, We conduct statistical research based on data from RoomView machine. We will use time series analysis to analyze the data which generated from RoomView system and run a SAS macro to process these data. We have interest to share this method so people can use that on similar tasks. Especially for machine learning environments, this method can perform a good data cleaning job. We will construct ARIMA models for ICT source: PC, DC, LA, CT, DVD and WPS. We will then compare these models using SAS macro with the previous research results from Williams and Xu (2014).

3   METHODOLOGY

3.1   Background

The public university used in this study is Radford University located in Radford, Virginia. The university was founded in 1910 as an all-women's college which became coeducational in 1972. Radford University currently has 180 electronic classrooms, each using a Crestron control system, spreading across 18 academic buildings on the campus. Each classroom has source components consisting of a dedicated computer (PC), a document camera (DC), a laptop connection (LA), a Blu-Ray/DVD/VCR combination player (DVD), a cable TV tuner (CT) and a wireless projection system (WPS). In addition to the source components, each classroom has an output component consisting of at least one digital projector and can have a secondary output component consisting of either a second digital projector and/or one to three TV monitors. Central to these components is the Crestron MPS-300 and a Crestron DM-8X8 which controls and routes the A/V signals

of the source components to the output component(s).

Beginning in October 2011 the University's IT department reprogramed the Crestron control systems to report data on how each classroom was used to the recording program, RoomView (version 7.2 also by Crestron). The data recorded by RoomView included in this study are the building and room number, the source component, and the date and time that a source component is used. By mid-January of 2012 sixty-eight of the Crestron controlled classrooms had been reprogrammed to send data to RoomView. The data collection continued through mid-September 2012 until RoomView was shut down to facilitate upgrading to a newer version. In late 2013, we decided to analyze the data which had been recorded to determine how the classrooms were being used and to alter any future classroom designs.

3.2   Raw data

The data set for our study consists of 68 classrooms. The earliest classrooms began recording data on October 13th 2011 and the last recording was on September 14th 2012. The data was processed using SAS 9.2 (32bit version) to divide the time that the equipment was used into 2 distinct time frames. The first time frame (referred to as "Prime Time") covers 7:30 a.m. through 7:30 p.m. Monday through Friday. "Prime Time" is when the majority of instruction occurs at the university. The second time frame (referred to as "After Time") covers 7:30 p.m. through 7:30 a.m. Monday through Friday and all day Saturday and Sunday. "After Time" is when the students are using the classrooms for studying and/or recreation. After processing, the data set was then converted to a time series format and exported for further analysis by R. The ARIMA models and plots of the usage of the equipment were constructed in R version 3.02.

Data0	Data1	Data2	Data3	Data4	Data5	Data6	Data7
10/26/2011 19:14	10/26/2011 23:35	source	PC	-	4.35	-	Cook Hall 107
10/26/2011 23:34	10/26/2011 23:35	source	CATV Receiver	-	0.016666	-	Cook Hall 107
10/27/2011 0:53	10/27/2011 0:54	source	PC	-	0.016666	-	Cook Hall 107
10/27/2011 0:55	10/27/2011 3:00	source	DVD	-	2.083333	-	Cook Hall 107
10/27/2011 12:50	10/27/2011 16:01	source	PC	-	3.183333	-	Cook Hall 107
10/27/2011 16:32	10/27/2011 17:46	source	PC	-	1.233333	-	Cook Hall 107
10/27/2011 20:41	10/27/2011 21:05	source	PC	-	0.4	-	Cook Hall 107
10/27/2011 21:05	10/27/2011 21:39	source	Laptop	-	0.566666	-	Cook Hall 107
10/27/2011 21:41	10/27/2011 23:12	source	PC	-	1.516666	-	Cook Hall 107

Figure 3.1 Raw Data

Firgure 3.1, Above, is a picture of the raw data exported into excel from the Roomview server. Data0 is the date-time stamp of when the system/source was started and Data1 is the date-time stamp of when the system/source was stopped. Data2, Data4, and Data6 are not used. We sorted the data by source type (Data3) to ensure that only the physical sources were displayed. When a specific location was required we would sort by the building/room number (Data7). Since our raw data set is nearly 30MB, we would then save the filtered data to a new spreadsheet to make working with the data we were interested in easier.

11/9/2011 15:01	11/9/2011 15:03	Source	CATV Receiver	-	0.033333	-	Young Hall 406
11/9/2011 15:01	11/9/2011 22:14	Source	VCR	-	7.216666	-	Young Hall 406
11/9/2011 15:02	11/9/2011 15:12	Source	VCR	-	0.166666	-	Young Hall 406
11/9/2011 15:03	11/9/2011 15:22	Source	VCR	-	0.316666	-	Young Hall 406
11/9/2011 15:03	11/9/2011 15:22	Source	VCR	-	0.316666	-	Young Hall 406
11/9/2011 15:07	11/9/2011 15:22	Source	VCR	-	0.25	-	Young Hall 406
11/9/2011 15:07	11/9/2011 15:12	Source	VCR	-	0.083333	-	Young Hall 406
11/9/2011 15:08	11/9/2011 21:08	Source	VCR	-	6	-	Young Hall 406
11/9/2011 15:08	11/9/2011 15:12	Source	DVD	-	0.066666	-	Young Hall 406
11/9/2011 15:10	11/9/2011 16:18	Source	VCR	-	1.133333	-	Young Hall 406
11/9/2011 15:10	11/9/2011 15:12	Source	CATV Receiver	-	0.033333	-	Young Hall 406
11/9/2011 15:10	11/9/2011 22:23	Source	VCR	-	7.216666	-	Young Hall 406
11/9/2011 15:12	11/9/2011 15:22	Source	VCR	-	0.166666	-	Young Hall 406
11/9/2011 15:12	11/9/2011 15:31	Source	VCR	-	0.316666	-	Young Hall 406
11/9/2011 15:12	11/9/2011 15:31	Source	VCR	-	0.316666	-	Young Hall 406
11/9/2011 15:16	11/9/2011 15:31	Source	VCR	-	0.25	-	Young Hall 406
11/9/2011 15:17	11/9/2011 15:22	Source	VCR	-	0.083333	-	Young Hall 406
11/9/2011 15:18	11/9/2011 21:18	Source	VCR	-	6	-	Young Hall 406
11/9/2011 15:18	11/9/2011 15:22	Source	DVD	-	0.066666	-	Young Hall 406
11/9/2011 15:19	11/9/2011 16:27	Source	VCR	-	1.133333	-	Young Hall 406

Figure 3.2 Snapshot of raw data from Young 406

Figure 3.2, above, is a snapshot of the raw data from classroom Young 406 that give us a large set of bad data. As

you can see from the snapshot there is a pattern for this room which repeats every few minutes. This room gave out bad data such as this for roughly 1 month, generating approximately 50 thousand data points which were removed before any analysis for this room was performed.

As we were converting the data to a Time series format we came across a classroom, cook 112, which was giving us an incorrect total number of minutes of usage in a day for a period from October 27th to October 31st 2011. The correct number of minutes of usage in a day should be no higher than 1440 for any given device in a day. No other classrooms that we could find showed this unusually high time usage. Upon inspection of the raw data (shown below) it was discovered that cook 112 also had a small batch of bad data similar to what was found with young 406. This data was also removed and all processing and analysis was rerun with this corrected data.

4   DATA PROCESSING

4.1   Sorting the data

| 10/14/2011 18:58 | 10/14/2011 20:08 | source | PC | - | 1.166666 | - | Young hall 207 |
| 10/14/2011 20:11 | 10/17/2011 10:55 | source | CATV Receiver | - | 62.733333 | - | Young hall 207 |

Figure 4.1 Raw data of Midnight problem

Figure 4.1, above, is an example of the raw data extracted from the room view server. The first cell in the example shows the date-time stamp of when the system was started or the source was shown, the second cell shows the date-time stamp of when the system was shut down or the source was switched away from. The third, fifth, and seventh cells contain no information relevant to this study. The fourth cell contains the source that was being displayed on the system. The sixth cell contains the number of hours that the source was displayed by the system. The eighth cell contains the building name and room number of the classroom.

Upon looking at the raw data, it was soon discovered that several data points had start and end dates that ended on different days. The second row of the above example shows this problem with the date starting around 9pm on the 14th of October 2011 and the end date around 11am on the 17th of October 2011, for a total of nearly 63 straight hours. Since our goal was to run a time-series analysis using daily data points, manipulation of the data set to correctly split the data points which crossed one or more midnights was necessary.

Most of the buildings around campus are open for students to use at roughly any given time. Since we currently do not have a method to determine if the electronic classrooms are being used by faculty or students, it was determined that the data for each date should be divided into two distinct time frames to best capture the usage by the respective groups. The first time frame (referred to as "prime time") would cover the majority of classroom instruction for the University, between 7:30 a.m. and 7:30 p.m Monday through Friday. The second time frame (referred to as "after time") would cover all other times that students would likely be using the classroom for studying and/or recreation, between 7:30 p.m. and 7:30 a.m. Monday through Friday and all day on Saturday and Sunday.

The final step for every data point is to be checked to see if it occurs during a weekend. If the data point does occur on a weekend the program will assign no time to "Prime_time" (PT) and assign all the time the source was used to "After_time" (AT). If the data point doesn't occur during the weekend the program then assigns the amount of "PT" and/or "AT" according to our definition of "Prime_Time"

In order to correctly separate the data into the final times we described in the above paragraph, we ran the data through a series of if-else statements. The first of the if-else statements was a check to see if the data point being read crosses midnight by looking at the variable "e". Variable "e" is an integer found using a SAS function (Intck) to see how many days have passed between the start and end cells of the data point. For data points that do not cross midnight, variable "e" will be equal to zero and the SAS program will preform the final step to assign the "Prime_time" and "After_time".

For data points that do cross midnight, the SAS program checks to see how many days the data points span and creates a do-loop using variable "e" plus one (the plus one is so that the last day is included in the do-loops iterations). When the do-loop is entered, SAS checks to see how many times the do- loop has run and preforms the appropriate final step for the appropriate day in the iteration to each day after it crosses midnight until it reaches the midnight just before the end date.

After all of the data has been processed using the final step. SAS then creates a Total_time variable which is a sum of the Prime_time and After_time. One last sort preformed is to prepare the data to be outputted as the final value. This sort organizes the data by building, room number and source then outputs the data as a new data set is used in our research.

4.2   The RoomView7 SAS Macro Setting

To use our SAS macro for sorting out the times into "Prime Time" and "After Time" you must first import your raw data set, making sure to use the getnames function and setting the out of your imported data set to CDSI. The syntax

of our macro is RoomView7(CDSI, LB=, UB=, WK=); The dataset must be named CDSI, has to be entered. The variables LB and UB are the lower (LB) and upper (UB) bounds of what you want your Prime Time to be. LB and UB are both character variables and are entered in as 'HH:MM'. If you do not specify these variables, the default values are for 07:30(7:30 am) for LB and 19:30(7:30pm) for the upper bounds. The variable WK is also a character variable and is used if you want the macro to include the weekends in its calculations of Prime time and After time. The default value for WK is 'no' which means it will not count the weekends as prime time and will record 0 for this value. Entering anything other than 'no' will cause the macro to treat the weekends just as it does for weekdays and will split the Prime and After time according to the Prime time bounds.

4.3   SAS MACRO

```
options nonumber;
    PROC IMPORT DATAFILE= "D:\XXX\XXX.xlsx"
    OUT= CDSI (drop= var3 var5 var7 )
    DBMS=EXCEL REPLACE;
    range= "c1:j15";
    GETNAMES=YES;
    MIXED=no;
    SCANTEXT=YES;
    USEDATE=no;
    SCANTIME=no;
    RUN;
    /* remove duplicate observations*/
    proc sort data = CDSI nodupkey out=inprogress;
    by start;
    run;
    Data inprogress;
    set inprogress;
    /*variables needed for functions involving the crossing of midnight*/
    /*Test variables, will be commented out in macro */
    LB = '07:30';
    UB = '19:30';
    WK = 'no';
    /*pulling out the date part of the start and end date-time*/
    a = datepart(start);
    b = datepart(end);
    /*finding the interval between the start(a) and end(b) dates */
    e = intck('day',a,b);
    day_s = day(a);
    day_e = day(b);
    /*Getting the weekday, month, and year of the start(a) and end(b) dates*/
    day_ws = weekday(a);
    day_we = weekday(b);
    month_s = month(a);
    month_e = month(b);
    year_s = year(a);
    year_e = year(b);
    /*variables needed for functions involving source time assignments*/
    /*Pulling the time part of the start and end date-time */
    c = timepart(start);
    d = timepart(end);
    /*Geting the interval, in minutes, of the time frames of intrest*/
    at1 = intck('minute','00:00:00't,'07:30't);
    at2 = intck('minute','19:30't,'24:00:00't);
    pt = intck('minute','07:30't,'19:30't);
    /*Separating building and room numbers along with the source used. */
    /*Finding the "Hall" part of the building name*/
    h_space = indexw(upcase(room),'HALL');
    /*Grabbing the Building name by looking in front of the "Hall" in the room variable */
    building = substr(upcase(room),1,h_space-1);
    /*Grabbing the Room number by looking after the "Hall" part of the room variable */
    room_num = upcase(substr(room,length(room)-2,3));
```

```
/*Grabbing the Source from the source variable */
Source_f = source;
i = 1;
/*For data not crossing midnight */
if e=
then do;
/*Check for Sunday and Saturday and to check to see if the weekend is included in the counting of prime time. */
if ((weekday(a) = 1) or (weekday(a) = 7)) and ((WK='no') or (WK='No') or (WK='NO'))
then do;
pt_f = ;
at_f = intck('minute',c,d);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
/*For days other than Saturday or sunday*/
Else do;
If(c <'7:30't and d <'7:30't) or (c >'19:30't and d >'19:30't)
Then do;
pt_f = ;
at_f = intck('minute',start,end);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
Else if ('7:30't <= c <= '19:30't) and ( '7:30't <= d <= '19:30't )
Then do;
pt_f = intck('minute',start,end);
at_f =;
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
Else if (c <'7:30't) and ( '7:30't <= d <= '19:30't)
Then do;
at_f = intck('minute', c,'7:30:00't);
pt_f = intck('minute','7:30't,d);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
Else if ('7:30't <= c <= '19:30't) and (d >'19:30't)
Then do;
at_f = intck('minute','19:30:00't,d);
pt_f = intck('minute',c, '19:30't);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
Else if (c <'7:30't) and (d >'19:30't)
then do;
```

```
temp = intck('minute',c, '7:30:00't);
at_f = (intck('minute','19:30:00't,d) + temp);
pt_f = intck('minute','7:30't,'19:30't);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
end;
/* */
end;
/*For data crosing Midnight*/
Else do;
do x = 1 to (e+1);
if x = 1
then do;
if ((weekday(a) = 1) or (weekday(a) = 7)) and ((WK='no') or (WK='No') or (WK='NO'))
then do;
pt_f = ;
at_f = intck('minute',c,d);
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
else do;
if (c <'7:30't)
then do;
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
at_f= intck('minute',c,'7:30:00't) + at2;
pt_f = pt;
end;
else if ('7:30't <= c <= '19:30't)
then do;
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
at_f= at2;
pt_f= intck('minute',c,'19:30't);
end;
else if (c >'19:30't)
then do;
at_f = intck('minute',c,'24:00:00't);
pt_f = ;
day_f = day_s;
day_wf = day_ws;
month_f = month_s;
year_f = year_s;
date_f = a;
end;
end;
output;
end;
```

```
Else if x = (e+1)
then do;
if ((weekday(b) = 1) or (weekday(b) = 7)) and ((WK='no') or (WK='No') or (WK='NO'))
then do;
day_f = day_e;
day_wf = day_we;
month_f = month_e;
year_f = year_e;
date_f = b;
at_f= intck('minute','00:00:00't,d);
pt_f= ;
end;
else do;
if (d <'7:30't)
then do;
day_f = day_e;
day_wf = day_we;
month_f = month_e;
year_f = year_e;
date_f = b;
at_f= intck('minute','00:00:00't,d);
pt_f= ;
end;
else if ( '7:30't <= d <= '19:30't)
then do;
day_f = day_e;
day_wf = day_we;
month_f = month_e;
year_f = year_e;
date_f = b;
at_f= at1;
pt_f= intck('minute','7:30't, d);
end;
else if (d >'19:30't)
then do;
day_f = day_e;
day_wf = day_we;
month_f = month_e;
year_f = year_e;
date_f = b;
at_f= at1 + intck('minute','19:30:00't,d);
pt_f= pt;
end;
end;
end;
Else do;
if (weekday(a + i) = 1) or (weekday(a + i) = 7) and ((WK='no') or (WK='No') or (WK='NO'))
then do;
pt_f = ;
at_f = at1 + at2 + pt;
day_f = day_s + i;
day_wf = weekday(a + i);;
month_f = month_s;
year_f = year_s;
date_f = a + i;
i = i+1;
end;
else do;
pt_f = pt;
at_f = at1+at2;
day_f = day_s + i;
```

```
    day_wf = weekday(a + i);;
    month_f = month_s;
    year_f = year_s;
    date_f = a + i;
    i = i+1;
    end;
    output;
    end;
    end;
    end;
    output;
    format date_f mmddyy10.;
    return;
    run;
    data inprogress;
    set inprogress;
    total_f = pt_f + at_f;
    output;
    run;
    /* Trouble shooting print out */
    proc print data = inprogress ;
    var building room_num source_f date_f day_wf pt_f at_f total_f;
    run;
    quit;
```

5 STATISTICAL ANALYSES

5.1 ARIMA Modeling of each Device

ARIMA is used here for the modeling and forecasting data. We will use the same time series modeling procedures to generate our results and compare it with previous study. ARMIA with no significant seasonal effects is tested and selected as preliminary result. We will use ARIMA model without seasonal effects, given by Equation 1.

$$\left(1 - \sum_{i=1}^{p} \Phi_i L^i\right) (1 - L)^d Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{1}$$

$$d \in Z \quad d > 0$$

Where L is the lag operator
and

$$\varepsilon_t = Y_t - \hat{Y}_t$$

is the forecast residual for a given time period.

We will select the best possible ARIMA model with p and q with a range of $0 \leq p \leq 5$ and $0 \leq q \leq 5$.

For the PC source component, the model which best fit the data was found to be model (4, 1, 5), given by Equation 2.

$$Y_t - 1.79Y_{t-1} + 2.18Y_{t-2} - 2.21Y_{t-3} + 1.9Y_{t-4} - .96Y_{t-5} =$$
$$\varepsilon_t - 1.08\varepsilon_{t-1} + 1.47\varepsilon_{t-2} - 1.13\varepsilon_{t-3} + .86\varepsilon_{t-4} - 3.12\varepsilon_{t-5} \tag{2}$$

For the Document Camera source component, the model which best fit the data was found to be model (4, 1, 5), given by Equation 3.

$$Y_t - 1.81Y_{t-1} + 2.21Y_{t-2} - 2.16Y_{t-3} + 1.71Y_{t-4} - .97Y_{t-5} = \varepsilon_t$$
$$-1.06\varepsilon_{t-1} + 1.41\varepsilon_{t-2} - 1.13\varepsilon_{t-3} + .81\varepsilon_{t-4} - .36\varepsilon_{t-5} \tag{3}$$

For the Laptop source component, the model which best fit the data was found to be model (2, 1, 1), given by Equation 4.

$$Y_t - 1.69Y_{t-1} + 1.05Y_{t-2} - 0.39Y_{t-3} = \varepsilon_t - 0.87\varepsilon_{t-1} \tag{4}$$

For the DVD source component, the model which best fit the data was found to be model (3, 1, 4), given by Equation 5.

$$Y_t - 1.37Y_{t-1} + 0.25Y_{t-2} + 0.99Y_{t-3} - 0.83Y_{t-4} =$$
$$\varepsilon_t - 1.05\varepsilon_{t-1} - 0.12\varepsilon_{t-2} + 1.23\varepsilon_{t-3} - .81\varepsilon_{t-4} \tag{5}$$

For the Cable TV source component, the model which best fit the data was found to be model (0, 1, 1), given by Equation 6.

$$Y_t - Y_{t-1} = \varepsilon_t - 0.91\varepsilon_{t-1} \tag{6}$$

For the WPS source component, the model which best fit the data was found to be model (3, 1, 3), given by Equation 7.

$$Y_t - 1.79Y_{t-1} + 1.21Y_{t-2} + 0.021Y_{t-3} - 0.42Y_{t-4} = \\ \varepsilon_t - 1.83\varepsilon_{t-1} + 1.79\varepsilon_{t-2} - 0.69\varepsilon_{t-3}$$

(7)

From these models (equation 2 to equation 7), we can see that we got almost identical models with SAS macro method and traditional data clearing method. Our observation attests the quality of the SAS macro. We have similar finding with the previous study. We can tell the PC, Document Camera, and Laptop connection usage is much more than DVD/VCR player, Cable TV, and WPS devices on campus.
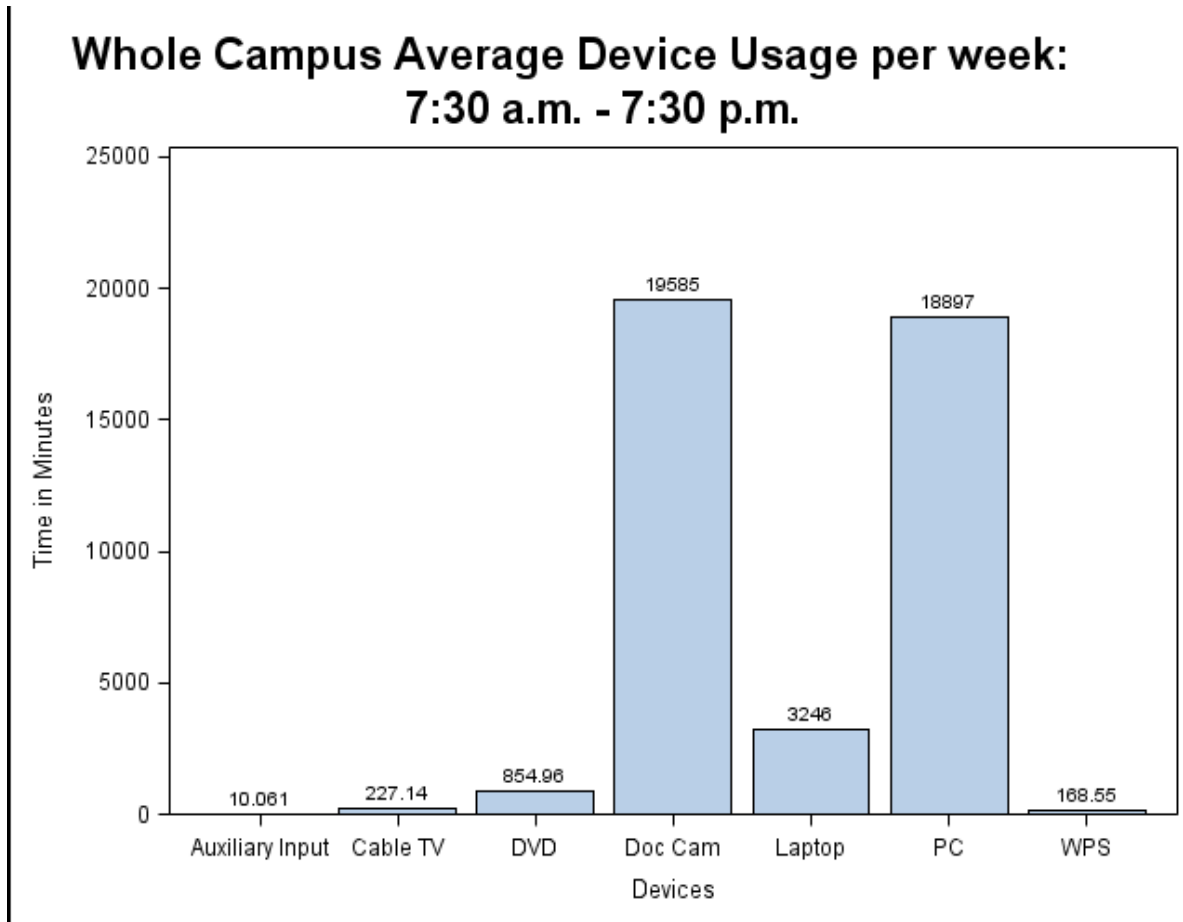

Figure 5.1 Device Usage during "Prime Time" for the Whole Camp

Similar results can be observed directly from Figure 5.1. This picture shows the whole campus device usage during prime time (office hours). We can see that the PC and Document Camera are leading the usage. Then we can observe Laptop, DVD, CT and WPS are the less used devices on campus.

6   CONCLUSIONS

In this study, we will mainly focus on how to handle the raw data from machine and how to utilize a SAS macro so that the work can become automatic. The macro is written in SAS and it is pretty robust. With little modifications, people can utilize this Macro to generate their own customized program to help them understand the data and eventually be able to do machine learning work.

From looking at the chart that have been generated for this data set, we can clearly see that the PC, Document Camera, and Laptop connections were the top three used devices on campus during prime time. Then we see the less usage of the following devices: DVD/VCR player, Cable TV, and WPS. The time series results attest to the fact that the DVD/VCR player, Cable TV, and WPS were used the least just like previous study showed. Judging from these results we should consider elimination of these least used devices from future electronic classroom designs.

The next step will be study of the none-prime time. We want to know the pattern of these devices. On the contrary of this study, if some device is used more during off peak time, it means people might just forget to turn them off. To better

understand the behavior of these devices, we can suggest either eremite these device or add auto shut off functions to these device to save electricity and further save energy in general.

REFERENCES

[1] Abdullah, L. (2012). ARIMA Model for Gold Bullion Coin Selling Prices Forecasting, *International Journal of Advances in Applied Sciences* (IJAAS),Vol. 1, No. 4, pp. 153-158

[2] Bowerman B.L., O'Connell, R. T. , & Koehler, A.B. (2005). Forecasting, *Time Series and Regression:An Applied Approach*. Thomson Brooks/Cole, Belmont, CA

[3] Brill, J. M. & Galloway, C. (2007). Perils and promises: University instructors' integration of technology in classroom-based practice. *British Journal of Educational Technology*, Vol 38 No 1, 95-105

[4] Choi, C., Esposito, C., Wang, H. X., Liu Z. & Choi J. (2018) Intelligent Power Equipment Management Based on Distributed Context-Aware Inference in Smart Cities, *IEEE Communications Magazine*, Volume: 56 , Issue: 7

[5] Chow, H. W. & Wu D. R. (2019) Outdoor Fitness Equipment Usage Behaviors in Natural Settings, *Int. J. Environ. Res. Public Health*, 16(4), 391

[6] Gülbahar, Y. (2008). ICT Usage in Higher Education: A Case Study on Preservice Teachers and Instructors. *The Turkish Online Journal of Educational Technology*, 1303-6521 volume 7 Issue 1

[7] Hyndman, R. J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software*, July 2008, Volume 27, Issue 3

[8] Roberson, J. A., Webber, C. A., McWhinney, M. C., Brown, R. E., Pinckard, M. J., & Busch, J. F. (2004). After-hours power status of office equipment and energy use of miscellaneous plug-load equipment. Report number LBNL-53729.

[9] Shenstone, L. & Hyndman, R. J. (2005). Stochastic Models Underlying Croston's Method for Intermittent Demand Forecasting. *Journal of Forecasting*, 24, 389-402.

[10] Shih, S. H. & Tsokos, C. P. (2008). A temperature forecasting model for the continental United States. *Neural, Parallel Sci. Comput*. 16, 1, 59-72.

[11] Shih, S. H. & Tsokos, C. P. (2008). Prediction models for carbon dioxide emissions and the atmosphere. *Neural, Parallel Sci. Comput*. 16, 1, 165-178.

[12] Shih, S. H. & Tsokos, C. P. (2009). A new forecasting model for nonstationary environmental data. *Nonlinear Analysis, Theory, Methods and Applications*, Vol. 71, No. 12, pp. e1209-e1214

[13] Usluel, Y. K., Aşkar, P. & Baş, T. (2008). A Structural Equation Model for ICT Usage in Higher Education. *Educational Technology & Society*, 11 (2), 262-273

[14] Vasileiadou, E. and Vliegenthart, R., (2013). Studying dynamic social processes with ARIMA modelling, *International Journal for Social Research Methodology*,
DOI: 10.1080/13645579.2013.81625

[15] Webber, C. A., Roberson, J.A. , McWhinney, M. C., Brown, R. E., Pinckard, M. J. & Busch, J. F. (2005). After-hours Power Status of Office Equipment in the USA. Report number LBNL-57470.

[16] Williams, J.& Xu, Y. (2014). Time Series Analysis for Device Usage in Electronic Classrooms. *JSM Proceedings, Boston, MA: American Statistical Association*, 3917-3929

[17] Xu, Y. and Choi, B., (2018). Analysis Seers Breast Cancer Data Using Random
Forest, *International Journal of Mathematical Sciences in Medicine(IJMSM)* , Vol 1, No2, Page 23–32