A Statistical Analysis of UA Wi-Fi Network Speed

For

STAT 571B Course project

(Prof.Lingling An)

By

Kensaku Okada

May 3rd, 2017

**Objective:**

The main objective of this project is to determine whether **location**, **time**, **day** and **PC (personal computers)** have any effect on UA Wi-Fi network speed. Specifically, we tried to answer the interesting questions like:

* Does location affect the UA Wi-Fi speed?
* Is the speed higher in mornings or evenings?
* Do the difference of days (weekdays and weekends) affect the network speed?

**UA Wi-Fi Architecture & Background:**

The UA Wi-Fi Network is a vast and comprehensive system with many components spread across the campus. The core component of the network is comprised of **3 Cisco routers**. These 3 core routers provide access to the World Wide Web for The University of Arizona. The core routers are “backed-up” by several **distribution routers**. In other words, when a core router fails, the distribution routers will provide the services without any network connection error. The next level down in the network infrastructure is switches. **Switches** serve to connect physical computers/devices to the routers and provide access to the network (a wired connection). Finally, **access points** connect to switches and allow for devices to connect to the network wirelessly. Our campus is said to have roughly 9400 access points on campus.

Coming to the device wireless network standards, **IEEE 802.11** Wi-Fi standards generally refer to the standard of the network module used in the device. There are a wide range of standards ranging from 802.11A/B/G/N/AC, and UA Wi-Fi can adapt to all of these specifications

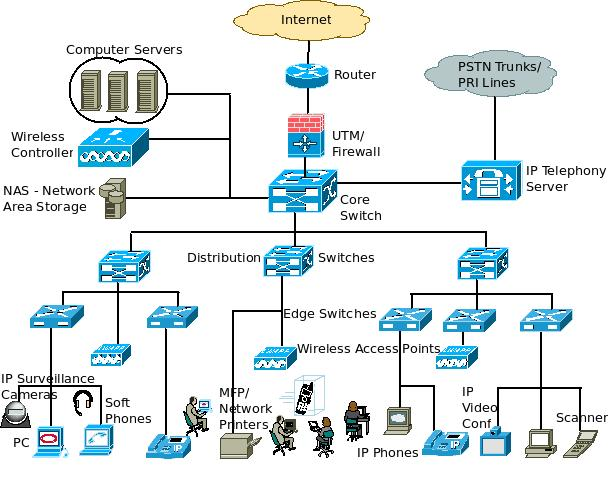


Figure 1. UA Wi-Fi Network Architecture

**Data collection & Methodology:**

To start the experiment, we wanted to know the overview and general functioning of UA Wi-Fi. We scheduled a meeting with the UITS team to gain an understanding of the overview of UA Wi-Fi network as well as effective methodology and experiment design. This meeting was very helpful towards our experiment and data collection strategy. Based on their suggestions, we were adopted a network speed checker tool called **iperf (**[**https://iperf.fr/iperf-download.php**](https://iperf.fr/iperf-download.php)**)** with which we were able to measure the absolute speed of the network. By using this software, we were able to successfully set up a virtual server and client system so that we can measure the network speed by-passing the internet having uncontrollable uncertainties such as network traffic, change of path and uncertain server locations.

In Figure 1, the actual communication path in our methodology was shown. The circled components acts either as a server or client, and they directly communication with each other.

Here is the detail of the methodology. We installed the iperf to each laptop. We fixed the location of the server at Science and Engineering Library which was connected with the network through an Ethernet port (not Wi-Fi) and measured the time at the three other locations. To eliminate as many nuisance factors as possible, the place of each location, position, direction and battery charge remaining were fixed at any combinations of treatment factors.

Hence, we succeeded to eliminate many nuisance factors by elaborating the design of experiment; first, by using iperf, we eliminated nuisance factors such as network path, unknown server locations, and other network traffic, etc, and second, by fixing the way of measurement, we also eliminated other nuisance factors like distance to the closest access point and existence of obstacles.

Based on our objective, as treatment factors, we chose three places, two time slots, and two days, and two different computers as shown below, and measured the network speeds at all combinations of them. The detail is described in the next section “Experimental Design”.

**Design of Experiment:**

In this section, we briefly explain the detail of the experimental design such as treatment factors, replications, method of statistical analysis.

We decided to do two experiments: a full factorial design, and a partial factorial design. The first design examined the interactions between all of the treatment factors below. In total, we had 24 combinations of the treatment factors.

1st Treatment- *Location*, 3 levels: Forbes Basement (1), Main Library (2), Shantz 440(3)

2nd Treatment- *Day*, 2 levels: Friday (1) & Sunday (2)

3rd Treatment- *Time*, 2 levels: 9 am (1) & 4 pm (2)

4th Treatment- *Who’s PC?.* 2 levels: Goutham’s (1) & Cameron’s (2)

One of the factors Day is usually considered to be a block factor in general, but we were interested in the interaction with Day, and its affect was also in the scope of our objective. Thus, we treated Day as a treatment factor. We collected 40 replicates for each measurement at a specific location, day and time. Hence we have 960 observations for the total experiment.

In the second experiment, we focused on the difference caused by the distance from access points and obstacles, and so, we also measured, only on Sunday, the speed at the connecting corridor at Shantz 4th floor, and compared the speed with that of Shantz 440. Same as the first experiment, we collected 40 replicates at each replication.

Lastly, the unit of response variable is Mega bits/second (Mbit/sec).

**Experiment 1 Results and discussions:**

The result of overall result of factorial analysis (Figure 2), LS-mean square plot for all 24 combinations of treatments (Figure 3), the boxplots for each treatment factor (Figure 4 to 7), and finally, result of normality test (Figure 8 and 9) and constant variance checking (Figure 10) were shown below. The more detail of other SAS analysis results is available in “analysis1” folder in “SASAnalysisResult” folder (you can see all results just by opening “sashtml4.htm” on browsers) .



Figure 2.Overall ANOVA and P-value table

From Figure 2, it was shown that the factor time has no significant effect to the network speed. We also saw that Day factor had a relatively larger p-value but still was considered statistically significant amongst the remaining factors and interactions.

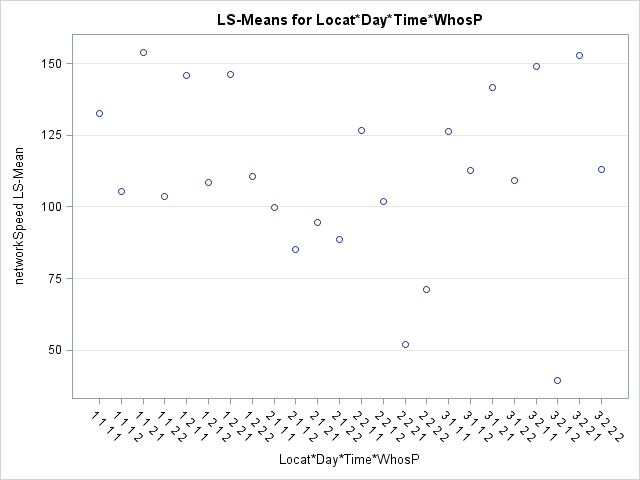


Figure 3.Least square means plot for all the combinations of treatment factors

Figure 3 shows the mean of response variables of all combinations of treatment factors. which marked 3 key remarkable outliers (surrounded by red and green). The red outliers were seen at (2,2,2,1) and (2,2,2,2,) (Main Library, Sunday. 4 pm with both computers). It indicates that the network speed at that state was especially slower than the other states.

The third key point was observed at (3,2,1,2) (Shantz 440, Sunday, 9am with Cameron’s PC). The reason could be because some computers at Shantz 4th floor ran a heavy network transaction with UA-Wi-Fi if we correctly measured the time without human error.

|  |  |
| --- | --- |
|  |  |
| Figure 4.Box plot of Time and Network speed | Figure 5.Box plot of PC and Network speed |
|  |  |
| Figure 6. Box plot of Day and Network speed | Figure 7. Box plot of Location and Network speed |

The series of box plots seen above (Figure 4 to 7) provides ranges, median values and mean values for network speed based on the treatment factors and their subsequent levels. From these plots, we can graphically confirm that the network speed was not so different at different Time and Day.

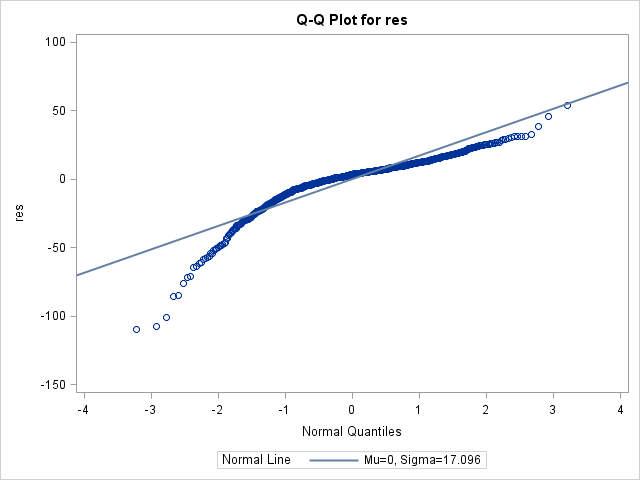


Figure 8. Q-Q plot for normality testing

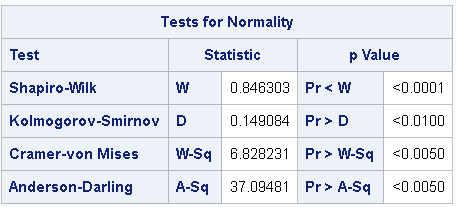


Figure 9. four numerical analyses for normality testing

From Figure 8, some data points apart from the normality line were observed. However the numerical analyses at Figure 9 indicated that the distribution was normal. I would be judged as normal because of a large number of replicates. In summary, it would be fine to regard the data normally distributes.



Figure 10. Plot for Variance checking

Like Figure 8, there were some data points off the range of residual in Figure 10, but most of the data were aggregated in the range of -50 to 30 residual. Thus, we would be able to deem the variance is constant.

**Experiment 2 Results and discussions:**

Because of the limitation of pages, we mentioned only the especially remarkable result of the experiment (please see “analysis2” folder in “SASAnalysisResult” folder for the result of the factorial analysis, the boxplots, normality test and constant variance test. You can see all results just by opening “sashtml.htm” on browsers).

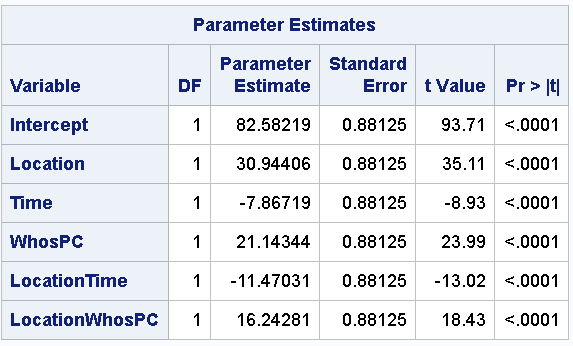


Figure 11. The result of regression analysis with main and interaction factors

Figure 11 indicated that the main factor Time has a significant difference between the Shantz connecting corridor and Shantz 440 on Sunday, which contradicted with the result of the first Experiment. It may be caused by the biased distribution of Time variance; it is possible that, in Experiment 1, the variance of time was diluted by taking the average of all combinations of treatment factors, but there might be some factors having quite large Time variance (we might see the result showing the significant difference at Time if we focus on specific locations even in Experiment 1). Further research, will be necessary for more precise discussion.

The result of Parameter Estimates showed that the main factor Location and WhosPC had large coefficient, which indicates these two factors have especially strong influence to the network speed. This discussion also agreed with the result of the first experiment (see the Mean Square at Figure 2); the mean squares of Location and WhosPC were much higher than the other main factors.

**Conclusion:**

In concluding our experimental statistical analysis of UA Wi-Fi network speed, it can be said that all interactions of treatment factors and 3 of 4 actual treatment factors (Location, Day, and types of computers) were statistically significant. The only insignificant treatment factor was Time. Among the 3 levels for ‘Location’, the Main Library provided the overall slowest network speed, while Forbes basement provided readings slightly greater than that of Shantz 440. Overall, ‘Time’ did not have much influence on network speed; “Day” was found to be statistically significant but did not seem to be so influential as the Location, types of computers and the other interactions.

As causes of the difference between the PCs used for measurement, many reasons would be come up with. To the best of our knowledge, the operation of the other application software, types of wireless card, and the compatibility between computers and UA Wi-Fi (more specifically, the access points) were suggested

Through this experiment a greater understanding of the UA Wi-Fi network infrastructure was achieved and data was successfully collected and analyzed. Finally, it can be said that the developed statistical model was both normal and adequate with almost 1000 data points collected.

**References and Thanks:**

Special thanks to: UITS team, Dr. Lingling An, and Brian Little

<http://www.ciscopress.com/articles/article.asp?p=25188&seqNum=4>

<http://www.radio-electronics.com/info/wireless/wi-fi/ieee-802-11-standards-tutorial.php>

**Appendix:**

**SAS code**

/\*Analysis 1 design of experiment\*/

/\*4 factorial design (levels:location, day, time, whosPC == 3,2,2,2)\*/

/\*replicates: 40\*/

/\*import data\*/

**data** projectAnalysis1\_data;

/\*change the following path wit your own path \*/

infile 'C:\Users\kensaku\_lenovo\Dropbox\universityOfArizona\STAT571BDesignOfExperiments\project\ProjectAnalysis1.csv'

/\*deliter == tab\*/

delimiter='09'x;

input Location Day Time WhosPC networkSpeed @@;

**run**;

**quit**;

/\*draw the data table\*/

**proc** **print** data=projectAnalysis1\_data;

**run**;

**quit**;

/\*referred to topic12\_factorial\_II.pdf p34\*/

**proc** **glm** data = projectAnalysis1\_data;

class Location Day Time WhosPC;

model networkSpeed = Location|Day|Time|WhosPC;

output out = projectAnalysis1\_data\_anova r=res p= pred;

**run**;

**quit**;

/\*multiple comparison. referred to topic12\_factorial\_II.pdf p13\*/

**proc** **glm** data=projectAnalysis1\_data;

class Location Day Time WhosPC;

model networkSpeed = Location|Day|Time|WhosPC;

means Location|Day|Time|WhosPC /tukey lines;

lsmeans Location|Day|Time|WhosPC/tdiff adjust=tukey;

**run**;

**quit**;

/\*3\*2\*2\*2 factorial regression analysis\*/

**proc** **reg** data = projectAnalysis1\_data;

model networkSpeed = Location Day Time WhosPC;

**run**;

**quit**;

/\*distribution normality checking\*/

/\*draw a qq plot \*/

title "normality checking";

**proc** **univariate** data= projectAnalysis1\_data\_anova normal;

var res;

/\*do not change this line below\*/

qqplot res/normal(mu=est sigma=est color=red L=**1**);

**run**;

**quit**;

/\*draw scattered plot for constant variance checking\*/

title 'residual plot: constant variance checking';

**proc** **sgplot** data= projectAnalysis1\_data\_anova;

scatter x=pred y=res;

refline **0**;

**run**;

**quit**;

/\*######################################################################\*/

/\*Analysis 2 design of experiment on sunday for shantz corridor and shantz 440 \*/

/\*2^3 factorial design (levels:location, day, time, whosPC == 2,1,2,2)\*/

/\*num of replicates: 40\*/

/\*import data\*/

**data** projectAnalysis2\_data;

/\*change the following path wit your own path \*/

infile 'C:\Users\kensaku\_lenovo\Dropbox\universityOfArizona\STAT571BDesignOfExperiments\project\ProjectAnalysis2.csv'

/\*deliter == tab\*/

delimiter='09'x;

input Location Day Time WhosPC networkSpeed @@;

/\*TimeWhosPC = Time \* WhosPC;\*/

LocationWhosPC = Location \* WhosPC;

LocationTime = Location \* Time;

**run**;

**quit**;

/\*2^3 factorial analysiss by regression analysis\*/

**proc** **reg** data = projectAnalysis2\_data;

model networkSpeed = Location Time WhosPC LocationTime LocationWhosPC ;

**run**;

**quit**;

/\*2^3 factorial analysiss by glm (ANOVA analysis)\*/

**proc** **glm** data = projectAnalysis2\_data;

class Location Time WhosPC;

model networkSpeed = Location|Time|WhosPC;

means Location|Time|WhosPC /tukey lines;

output out = projectAnalysis2\_data\_anova r=res p= pred;

**run**;

**quit**;

/\*distribution normality checking\*/

/\*draw a qq plot \*/

title "normality checking";

**proc** **univariate** data= projectAnalysis2\_data\_anova normal;

var res;

/\*do not change this line below\*/

qqplot res/normal(mu=est sigma=est color=red L=**1**);

**run**;

**quit**;

/\*draw scattered plot for constant variance checking\*/

title 'residual plot: constant variance checking';

**proc** **sgplot** data=projectAnalysis2\_data\_anova;

scatter x=pred y=res;

refline **0**;

**run**;

**quit**;