RED LIGHT RUNNING AT TRAFFIC CIRCLES: ESTIMATION AND EVALUATION OF COUNTERMEASURES EFFECTIVENESS

A Thesis presented to the Faculty of the Graduate School of the

University of Missouri – Columbia

In Partial Fulfillment of the Requirements for the Degree

Master of Science

By

BORIS RAUL CLAROS

Carlos Sun, Ph.D., P.E., J.D., Graduate Advisor

MAY 2013

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

RED LIGHT RUNNING AT TRAFFIC CIRCLES: ESTIMATION AND EVALUATION OF COUNTERMEASURES EFFECTIVENESS

presented by Boris Raul Claros, a candidate for the degree of Master of Science, and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Professor Carlos Sun, Ph.D., P.E., J.D

_____

Professor Praveen Edara, Ph.D., P.E.

_____

Professor John Fresen, Ph.D.

ACKNOWLEDGEMENTS

The research was conducted by Boris Raul Claros. A graduate student sponsored by the international Fulbright Scholarship Program at the University of Missouri.

The author would like to acknowledge the support and guidance provided by the academic advisor Dr. Carlos Sun, Associate Professor and Director of Graduate Studies of the Department of Civil and Environmental Engineering. Additionally, the assistance of undergraduate students Tyler Horn, Kevin Koines, Igor Caus, and Dallas Crain for their contribution in the video processing to obtain the dataset for this research is greatly acknowledged.

Finally, the unconditional support of close family members and friends is immensely appreciated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The complicated legacy design of traffic circles results in operational and safety deficiencies. This type of intersection design is commonly used at major roadways in Cochabamba, Bolivia. Traffic circles are intended to accommodate various types of vehicles and turning movements, and to store queues along the circulatory roadway, but the actual operational performance is poor.

Field observations, undergraduate research experience, and a review of current literature helped to identify the most critical traffic circle performance issue as driver compliance. This research focuses on the main issue of traffic signal violation at traffic circles, and specifically red light running.

The objective of this research is to analyze red light running at three intersections along the same corridor in Cochabamba, Bolivia. As a result of the analysis, different parameters are identified to verify the effectiveness of current local traffic management implementations, develop a red light running estimation model, and propose potential solutions.

Traffic data was collected over the summer of 2012. Video was recorded of traffic movements from main approaches and speeds at 200 meters upstream from the intersection to determine approach speeds. A video display along with the use of pre-defined keyboard keys is used to optimize the data processing. The data processing was performed by the author and four outstanding undergraduate civil engineering students. In order to produce uniform results, the students were trained on the same methodology before they processed the data. The analysis of the red-light-running data involved: traffic

flow, approaching speed, traffic signal control features, and geometric characteristics. A comparison among the different scenarios was used to determine if improvements were significant.

The results of this research were found to be statistically significant and were similar to previous studies. Initially, increase in red light running rates were found to relate to higher approaching speeds and higher traffic volumes. The implementation of count-down panels in the traffic signal control did reduce red light running, but on a short-time basis only. A geometric conversion from a traffic circle to a four-legged conventional intersection was found to be a poor countermeasure for safety and operational performance. Regarding the model for estimating red light running rates, the most significant variables included in the model were Y (yellow running rate), A (small vehicle traffic flow), and G (single unit truck traffic flow).

Four major recommendations resulted from this study. First of all, countdown panels are desirable despite losing effectiveness over time, since they have other benefits during congested scenarios. Second, an increase of the yellow signal interval would achieve significant reduction on red light running rates. Third, the implementation of an all red interval would contribute to the clearing of intersections and avoid potential incidents due to red light running. Finally, as a larger scale solution, it is desirable to separate the main roadway form the circulatory roadway by using an underpass to keep the traffic circle at ground level.

CHAPTER 1. INTRODUCTION

1.1. Overview

Traffic circles are old style circular intersections used in many cities of Bolivia where traffic signals might be used to control one or more entering points. As a result, traffic circles have distinctly different operational characteristics from modern roundabouts, with queue storage within the circulatory roadway and progression of signals. The inherit design does not provide any protective measure for pedestrians, and its overall operational safety measures are greatly questioned. The implementation of this particular intersection has been common in Cochabamba, Bolivia as the main vehicle distributor at major roads. There are several problems with the operational and safety performance of traffic circles at major corridors.

First of all, the traffic volume in the major roadway is considerably larger than the minor roadways. It generates a significant interruption of the flow because of the time spent in directing the different turning movements from or toward the minor road. Figure 1.1 shows an example of turning and through movement volumes from a major traffic circle at the same city (Claros and Gonzales 2008).

Second, in general the speed limit in this major roads is 70 (km/hr), but vehicles go beyond that limit, and the predominant speeds on the faster lanes ranges from 70-120 (km/hr). Furthermore, at traffic circles, there are not safety measures for pedestrians because stop and pedestrians pavement markings are not yielded by drivers, and people usually have to cross behind the first vehicle that completely stops at the approach.

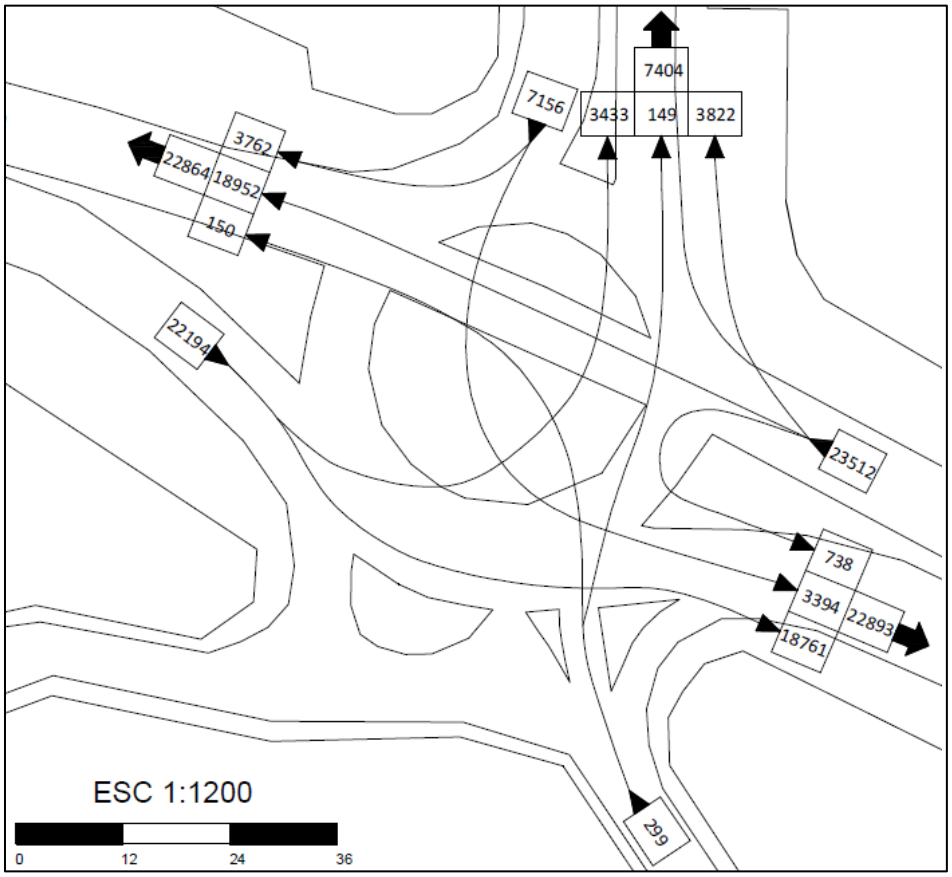


Figure 1.1. AADT 2008 ABC Traffic circle Cochabamba, Bolivia, Claros and Gonzales (2008).

However, there are several pedestrian bridges close to the intersections, they are not very popular, and very few people use them. Also, fence or barriers were built in the median to restrain people to the assigned facilities, but it did not last long because the fence was cut or the barrier was torn down.

Regarding vehicle accidents, the most concerning safety issue was the right angle crash record which outlines that 43.1% of the accidents at traffic circles were either from merging or right angle collisions as observed in Table 1.1 (Claros and Gonzales 2008). From field observations, it was identified that one of the most concerning problems regarding signalized traffic circles was the red light running during overall conditions,

and the excessive speed at which vehicles travel through the intersection during low flow conditions, vehicles didn't have to meet any queues (no congestion present).

Table 1.1. Accidents by crash types at traffic circles (Claros and Gonzales 2008).

N	Crash type	Accidents	%
1	Crossing or merging collision	94	43.1
2	Queuing or diverging collision	55	25.2
3	Loss of control	30	13.8
4	Pedestrian	17	7.8
5	Bicyclist	5	2.3
6	Motorcycle	13	6.0
7	Passing a bicycle	1	0.5
8	Passing a motorcycle	3	1.4
Total		218	100.0

The traffic signal operation usually worked well under low flow conditions, but it was troublesome at peak hours with high flow of turning movements. Since traffic circles allow storage of turning vehicles on the circulatory roadway, with high volumes, this area filled out fast, and queues developed in the major road restraining the trough movement to travel across the intersection at full capacity. Also, since under such congested conditions were present, drivers usually tried to merge with the groups of vehicles even after the traffic signal changed to red, they felt protected by the big platoon, so these took some time away to clear the intersection. In Figure 1.2, a visual example of a saturated condition at a traffic circle is displayed.

There is a wide range of potential countermeasures to the red light running problem. These solutions are generally divided into two broad categories: engineering countermeasures and enforcement countermeasures. Enforcement countermeasures are intended to encourage drivers to adhere to the traffic laws through the threat of citation or

possible fine. In contrast, engineering countermeasures are intended to reduce the chances of a driver being in a position where he or she must decide whether or not to run the red. Some studies (Bonneson and Zimmerman 2001) showed that countermeasures in both categories were effective in reducing red light running.

Figure 1.2. Congested conditions at a traffic circle.

1.2. Research Objectives

The research intents to evaluate red light running at the main approaches of traffic circles under different flow conditions, approaching speeds, vehicle mix, traffic signal characteristics, and associated case scenarios. These objectives were achieved by satisfying the following goals:

- Analysis of the predominant red light running rates.
- Analysis of speed on red light running occurrence.
- Comparisons among different intersections and cases.
- Before and after comparison of countdown panel implementation.
- Provide a multivariable linear regression model in order to facilitate a tool for local agencies to estimate red light running rates according to flow management decisions and countermeasures implementations.

- Identify promising countermeasures.

1.3. Research Scope

This research project deals exclusively with the evaluation of red light running at traffic circles, four legged intersection, a proposed regression model for prediction, and feasible countermeasures for the specific region of study.

1.4. Research Approach

The research approach is based on a four stage structure focused on organization, collection of the data, data processing, and data analysis. Initially, the organization focuses on determining the scope of the research considering the resources available. The selection of the locations of interest for the study was conducted following an elaborated schedule for the collection of the data. Following the data collection plan, during eight weeks, at specific locations and times of the day, video recordings of traffic and speed measurements were conducted. A video processing method was developed to determine flow rates with specific turning movements, vehicle classification, and signal violation. Also, with a similar video processing procedure, speed measurements were recorded according to vehicle classification and spatial location on the roadway. Finally, an analysis of the different variables involving red light running violations from the processed dataset were analyzed to build a regression model and propose potential countermeasures.

CHAPTER 2. LITERATURE REVIEW

2.1. Introduction

The literature review focuses on current studies analyzing red light running. First of all, a summary was conducted about research on specific analysis of contributing factors: flow rate, signal cycle, probability of stopping and yellow interval duration. Research regarding the different prediction approaches of the red light running was discussed. Finally, potential countermeasures and their performance on the field after implementation were also reviewed in this section.

2.2. Contributing Factors for Red Light running

Bonneson and Zimmerman (2001) classified contributing factors as "Exposure" and "Contributory" factors. Exposure factors represent the basic events that must occur or be present for a red light to be run. These factors are not necessarily considered to be causes of red light running. In contrast, contributory factors represent conditions or behaviors that cause or facilitate red light running.

2.2.1. Exposure Factors

The exposure factors are circumstances that set the stage for the contributory events that follow to precipitate the red light running. Some factors that underline these events include flow rate and signal cycle (Bonneson and Zimmerman 2001). Flow rate is an important measure to analyze red light running because the exposure of vehicles approaching the traffic control signal during yellow signal display potentially generates certain trend toward the red light running. It is likely that drivers running the red signal

during each cycle will likely increase as flow rate increases. In Figure 2.1, flow rate to cycle length ratio and frequency of red light running are used. The figure suggests that red light running frequency increases with increasing flow rate or shorter cycle lengths. The pattern indicates that the relationship is somewhat linear (Bonneson and Son 2003).

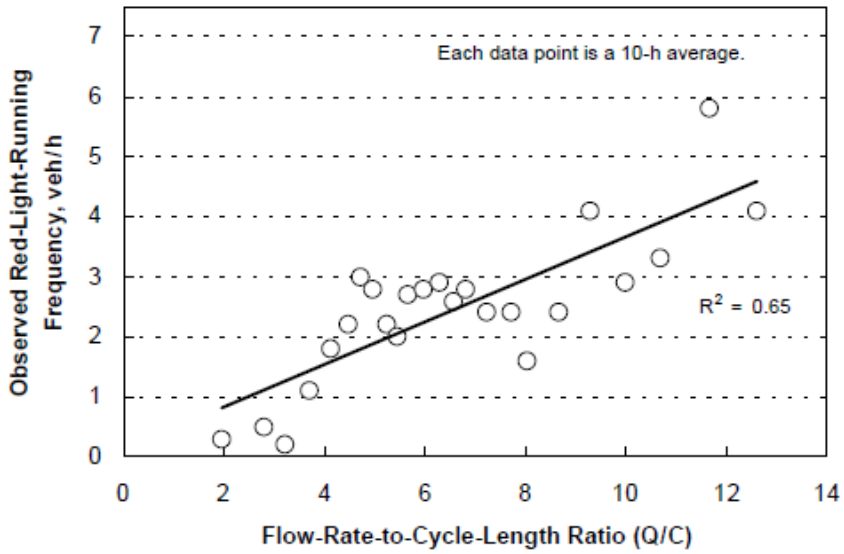


Figure 2.1. Red light running frequency as a function of flow rate to cycle ratio, Bonneson and Son (2003).

Similarly, Bonneson and Zimmerman (2001) combined the results of two different researches in one single graph as can be observed in Figure 2.2. It involves red light running as a function of approach volume at urban with no advanced detection and rural with advanced detection intersections.

A study conducted in Virginia, covered six traffic control intersections at three different cities. They determined that higher red light running rates were observed in cities with larger intersections and higher traffic volumes (Porter and England 2000). Research conducted at seven intersections in England concluded there is significant

correlation between red light running occurrence and the approach volume (Baguley 1988). Finally, in a safety regression analysis to determine potential locations for safety improvements, it is observed that red light running crash frequency is correlated to the approach leg AADT (Mohamedshah et al. 2000).

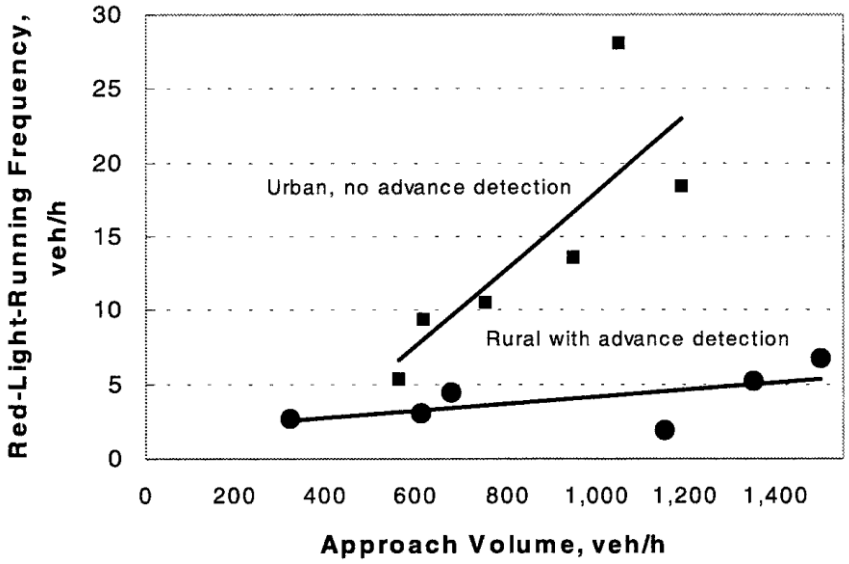


Figure 2.2. Red light running frequency as a function of approach volume, Bonneson and Zimmerman (2001).

2.2.2. Contributory Factors

The contributing factors are circumstances in which the conditions established or the driver's behavior leads red light signal violations. The most noticeable factors considered in red light running analysis and prediction models are the probability of stopping and the yellow interval duration (Bonneson and Son 2003).

From the exhaustive literature review performed by Bonneson and Zimmerman (2001), the common findings of this research was that the probability of stopping, when the yellow light was displayed, increases with the driver's travel distance to the

intersection. Additionally, there are other factors that play an important role in the probability of stopping. For instance, short headways associated with platoon, flow, control mode, approach grade, threat of crash, threat of citation, and expected delay (Bonneson and Son 2003).

A generally identified parameter in the red light running behavior is the yellow interval duration. Several studies suggest that the yellow interval should be based on the 85th (or 90th) percentile driver's travel time to the stop time. For instance, in Figure 2.3, a yellow interval of 4.2 seconds is sufficient for 85 percent of the drivers (Bonneson and Zimmerman 2001, Bonneson and Son 2003).

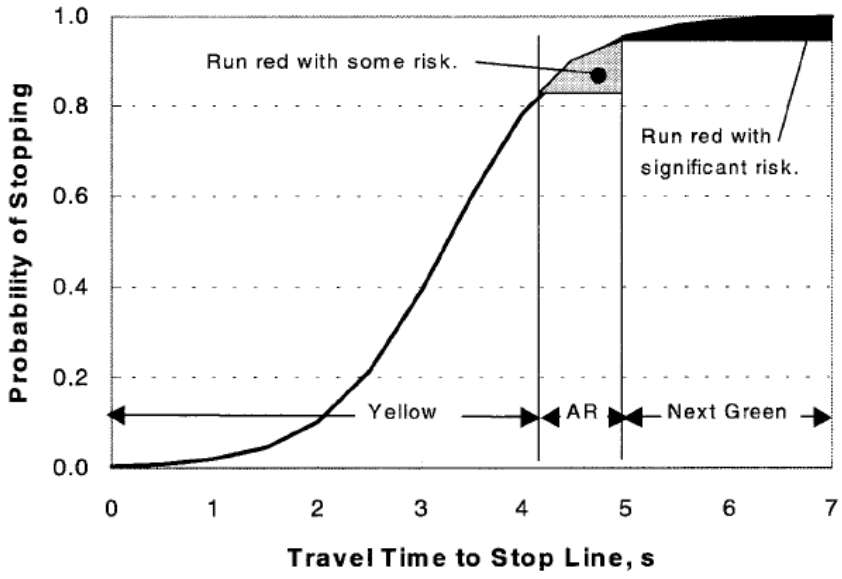


Figure 2.3. Relationship between probability of stopping and travel time to stop line, Bonneson and Zimmerman (2001).

2.3. Red Light Running Prediction Models

Different variables have been used in models to predict red light running behavior. The most common variables are the overall flow rate, cycle length, yellow phase, or percentage of heavy vehicles.

Bonneson and Zimmerman (2001) calibrated a regression model using the negative binomial approach to predict the frequency of red light running using phase-end flow rate, cycle length, and propensity of drivers on a given intersection to run the red light. Additionally, they incorporated the effect of the yellow interval duration and percentage of heavy vehicles. The study confirmed that the yellow interval duration (less than 3.5 seconds) had a significant effect on the red light violation. Also, the presence of heavy vehicles increased the number of vehicles running the red light, but less significantly. Finally, the analysis of the approach volume on the red light running frequency revealed a high correlation.

Multivariable linear regression and Poisson distribution methods have been commonly used for modeling count data. However, Yiyi et al. (2009) considered generalized estimating equations with the negative binomial as the function link. The study considered five intersections and individual turning movements to be modeled separately (protected left turning and through movements). The frequency of the red light running is found to be related to the traffic flows to which the conflictive traffic belongs. Large traffic flows of the conflicting traffic do not necessarily increase the frequency of the red light running. With the inclusion of geometric design features in the model, the number of approaching lanes, through lanes, and the speed limit compliance were

significant in the prediction of red light violation. Additionally, long crossing distance and yellow signal interval decreased the frequency of red light running.

A different approach involving propensity to red light violation has been developed by Giuffrè and Rinelli (2006). It was considered that the violation of the red light was a result of two main factors: when circumstances make it possible and the decision to stop at the intersection or keep going. These factors are human factors and operational characteristic present at the location. With a quantitative evaluation of red-light running, propensity was conducted using speed, yellow light duration, and clearance path length. The number of violation maneuvers is the result of individual decision processes in the face of operational circumstances that make the red light running violation possible. As a result, they consider that potential violators can be identified with a specific site. The specific cases studied suggest that road factors have little influence on the red light running violation when there were low traffic conditions. In contrast, with higher traffic volumes, diverse driver behavior was observed. With the results obtained, they consider that red light running rates can lead to incorrect evaluation of the violation because it is not suitable to represent propensity in this matter. Therefore, the effective number of violations depends on red light violation opportunities resulting from operational characteristics of intersections. The recommendation of the authors is to survey opportunities rather than only light violations rates to specific location analysis.

Alternatively, the focus of estimating the likelihood of the red light running was also determined using specific car locations on the roadway, interaction with other vehicles, site location features, and current conditions when the red light running occurs. Elmitiny et al. (2010) used the following variables to estimate the red light running: (1)

whether the vehicle was leading or following the traffic, (2) the vehicle position on the roadway (lane), (3) Vehicle type (passenger car, light truck vehicle, larger size vehicle), (4) vehicle's operating speed at the onset of yellow, and (5) the distance from the vehicle to the stop bar at the onset of yellow. The independent variable with more significance in the model was the distance. Drivers tend to cross the intersection as the yellow onset distances decreased. Drivers are more likely to run red lights at distances where it seems possible to beat the signal change. If the vehicles' yellow-onset distances are between 267.5 (ft) and 372.5 (ft), the operating speed plays an important role in the stop/go decision and red light running. More than 50% of speeding drivers would make go decisions at these yellow-onset distances. Also, more than 50% of speeding drivers would run red lights at the yellow-onset distance interval between 292.5 (ft) and 372.5 (ft), and these speeding red-light runners constitute a substantial portion of the overall red light running violations." (Elmitiny et al. 2010). Therefore, the effect of reducing the approaching speed plays a significant role in reducing the red light running violations. A significant finding from the study suggests that drivers in the following position were more likely to violate the signal than the leading drivers. "It was found from the analysis that drivers' stop/go decisions may be dependent of vehicle's yellow-onset distance, speed, and other traffic parameters and logistic regression assumes that a link function can be used to relate the probabilities of group membership to a linear function of the predictor variables." (Elmitiny et al. 2010).

2.4. Countermeasures

There are several countermeasures that can be used to reduce red-light running. There are two types: engineering and enforcement countermeasures. Engineering

countermeasures focus on reducing red light running violation frequency when drivers face a circumstance where they have to decide whether or not to run the signal. Enforcement countermeasures concentrate on compliance of driver to traffic laws through fines or any other penalization (Bonneson and Zimmerman 2001). This section of countermeasures focuses on engineering countermeasures.

According to Bonneson and Zimmerman (2001), there are two types of red light running drivers: intentional and unintentional. The intentional driver is the one that violates de signal in response to the excessive delays or congested conditions. In contrast, the unintentional driver is the individual that is not able to stop at the onset of the signal due to diverse circumstances of traffic control, geometry, visibility, approaching speed, or multiple combinations.

There are several engineering countermeasures that are applicable to specific scenarios; therefore, the most relevant were discussed. There are important contributions regarding the influence of the yellow time interval, all red interval, unwarranted traffic signals, and pre yellow information.

2.4.1. Increase of the Yellow Interval Duration

Increasing the length of the yellow phase interval contributes in reducing the red light running (Bonneson and Zimmerman 2004). An increase on one second, from 3 to 4 in urban areas and from 4 to 5 in rural areas, reduce red light running by 50% (Van der Horst and Wilmink 1986).

The most known and commonly used methods to determine the length of the yellow phase are the equations recommended in ITE guidelines (ITE Technical Council Task Force 4TF-1, 1994).

- Yellow interval

$$Y = t + \frac{v}{(2a \pm 2Gg)} \qquad \text{(E 2.1.)}$$

- Yellow interval (all red interval clearance not used)

$$Y = t + \frac{v}{(2a \pm 2Gg)} + \frac{(W+L)}{v} \qquad \text{(E 2.2.)}$$

Where:

t, driver perception–reaction time for stopping, taken as 1 s.

v, approach speed, ft/s (m/s), taken as 85th percentile speed.

a, deceleration rate for stopping, taken as 10 ft/s^2 (3.0 m/s^2).

g, percent grade, divided by 100.

G, acceleration due to gravity taken as 32.2 ft/s2 (9.8 m/s^2);

W, width of intersection, in ft (m), measured from the upstream stop bar to the downstream extended edge of pavement; and

L, length of clearing vehicle, taken as 20 ft (6.1 m).

Although these equations have been used for many years, several researchers suggested that the yellow interval duration should be based on the probability of stopping (Bonneson and Zimmerman 2004, Van der Horst and Wilmink 1986). Therefore, they recommend that the yellow interval should be based on the 85^{th} (90^{th}) percentile drivers' travel time to the stop line.

In order to assess the influence of changing the yellow interval, researchers have used before and after comparisons. Diverse results were found, but they suggest that the benefit of increasing the yellow interval does contribute to the reduction of red light running.

Bonneson and Zimmerman (2004), conducted a study involving 8 site locations in Texas for the before and after study of yellow interval duration. The after study was conducted 6 months after the before study. With 3370 signals cycles collected, they found that an increase of 1.0 second in the yellow phase (less than 5.5 seconds) decreased the frequency of red light running by at least 50%, which supports the results from previous research. Additionally, it was observed in the study that drivers adapt to the yellow interval changes which generate a slightly lower probability of stopping at the onset of yellow. The results of the study involving yellow signal interval change and red light running are displayed in Figure 2.4.

Schattler et al. (2003) also conducted a before and after study to determine the yellow interval influence on red light running. The study considers three intersections in Michigan that contained major highways intersected by low volume suburban arterials. The methodology conducted for the analysis consists on using the already functioning

yellow signal timing and adjust it according to the calculations and results with the ITE equations. Therefore, the yellow ranged from 0.1 to 0.6 seconds on the arterials, and 0.4 to 1.3 on the main approaches. The before and after study showed mixed results. One of the intersections recorded significant reduction of red light running when the yellow interval duration was increased from 0.1 to 0.4 seconds. However, at the other two intersections, no significant difference was found from the before and after changes.

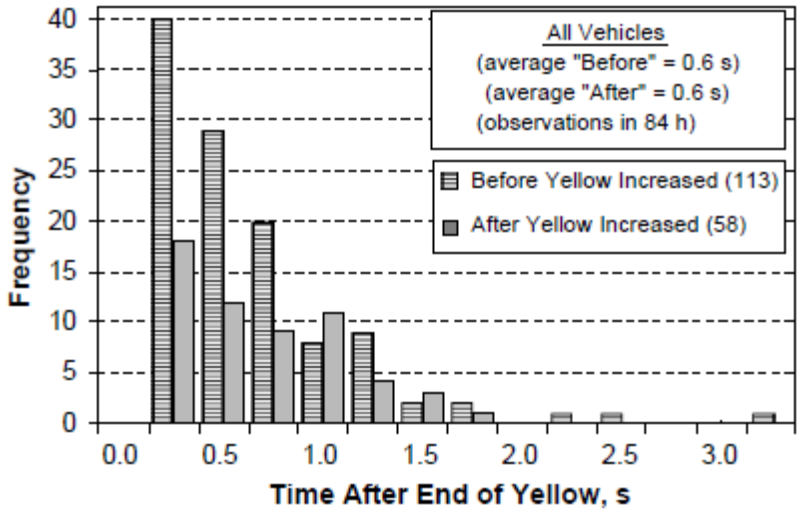


Figure 2.4. Effect of an increase in yellow-interval duration on the frequency of red-light running, Bonneson and Zimmerman (2004).

2.4.2. Increase of All Red Interval Duration

Although the all red interval plays a more important role clearing up the intersections and contributing to overall safety of operations, some research suggests its contribution on reducing red light running.

Schattler et al. (2003) in the same study of assessing the increase in the yellow interval duration simultaneously conducted the analysis of influence of all red light interval variation. The rates of late entry, involving violation of the signal, significantly

decreased after the increase of 2.0 to 3.0 seconds of the all red interval. Consequently, the contribution of increasing the all red interval was effective to reduce red light running and the risk of late exiting vehicles that could represent a safety issue.

Additionally, Bonneson and Zimmerman (2004) cited two studies in which there are contradictory results. The first study did not find any significant contribution of increasing the all red interval. On the other hand, the second article found that the nominal increase did actually reduce the frequency of right angle crashes by 40%, which were found to be potentially related to red light running behavior.

2.4.3. Removal of Unwarranted Traffic Signals

The removal of unwarranted traffic signals is a feasible countermeasure in low traffic volume conditions. Some studies found that it could significantly reduce the red light running and intersections crashes (Retting et al. 1998). A research conducted in Philadelphia also studied the effect of converting one way intersections from signal to multiway stop sign control. "Aggregate results indicate that replacing signals by multiway stop signs on one-way streets is associated with a reduction in crashes of approximately 24%, combining all severities, light conditions, and impact types." (Persaud et al. 1997).

2.4.4. Provide Pre-Yellow Information

Providing an advance warning signal to alert from yellow signal display is a common countermeasure to reduce red light running and overall safety at signalized intersections. Its effectiveness was tested, and some mixed results were found.

Grant et al. (2009) studied the performance of advanced warning signals in a longer monitoring basis: two years before and two years after. The advanced warning system contributed to maintain optimal operating speeds which contributed to capacity since vehicles did not need to worry about stopping. In contrast, speeds reduced in the time just before the onset of the yellow interval. Also, drivers tend to adjust to the warning signal, and they tried to beat the light. As a result, there was evidence of an increase of red light running even with shortening of the lead flash time. However, this increase did not contribute to the increase of overall crash frequency. Overall crash frequency was reduced 7% after installation.

Grant et al. (2007) also conducted a before and after installation of advanced warning signals. The study was conducted in Utah, collecting data prior to and immediately after installation. The results showed that during active warning of the system, the mean speeds reduced from 5 to 10 miles per hour. Additionally, the system performed well right after installation because a significant reduction on red light running rates is recorded. The reduction consisted of 5 violations per million entering vehicles to 1 violation per million entering vehicles.

Additionally, Bonneson and Zimmerman (2001) cited a study (Farraher et al. 1999) from the Minnesota Department of Transportation's use of advanced warning signals at high speed isolated intersections. The measurements indicate that the reduction on red light running after implementation of the system was 29%.

Consequently, multiple studies suggest significant contributions of implementing advanced warning signals to reduce red light running. However, some of the studies

suggest that the performance of the system was more significant right after the implementation, and its contribution reduced gradually in time due to driver's adaptability to the system. Also, its implementation seemed to be more effective in specific locations instead of adopting a widespread policy of implementing the system in several locations along the same corridor or region (Grant et al. 2007, Grant et al. 2009, and Bonneson and Zimmerman 2001).

2.4.5. Automated Systems and In Person Enforcement

Automated enforcement systems are among the most popular and effective countermeasures to red light running in the United States. It consists of digital camera and detector systems to capture photographs of vehicles violating the signal. The image serves as evidence to recognize the vehicle and identify the owner to mail the fine. An efficient and updated vehicle registration record is required to effectively run the system. Unfortunately, Bolivia has a poor registration information system and the applicability of red light cameras is not feasible.

In person enforcement is very popular in Bolivia since it can easily be deployed. It is commonly used during peak hours at highly congested intersections. However, the lack of work force does not allow providing the measure on a regular basis, and it is applied only when major events take place. Similarly, its applicability as a potential countermeasure to reduce red light running is limited.

CHAPTER 3. DATA COLLECTION

3.1. Introduction

This chapter focuses on describing the procedures followed to conduct the video recording, data processing, and analysis of red light running rates at traffic circles in Cochabamba, Bolivia. The intersections selected were three. They were chosen, because they contain particular traffic signal sequences, they were located at importance junctures, they had unique geometry, and they contained additional control devices. These three intersections were suitable for meeting the objectives of this research.

3.2. Description of Locations

As shown in Figure 3.1, the intersections are located one after the other in sequence along the main corridor (Blanco Galindo). They are named, for the purpose this study, according to the intersecting arterials. Blanco Galindo is a main corridor between the city of Cochabamba and the town of Quillacollo. The roadway has three lanes, nine signalized traffic circles, similar geometric features throughout, and fairly constant horizontal and vertical alignment. It has an AADT of 38000 (vehicles/day/direction), with diverse mix of vehicles. It is classified as the main transportation connector to the city of La Paz which is the country's capital. All the heavy vehicles between the two cities circulate through this facility, and the traffic circles are the locations that provide turning points where vehicles can deviate from the mayor roadway to different arterials. Public transportation vehicles are very significant on this corridor, since it is the main mode of transportation due to predominant social economic characteristics.

Figure 3.1. Location of intersections along main corridor.

3.2.1. PABLO Intersection

Pablo intersection is where one countermeasure was implemented. Initially the intersection had the same features as the other two intersections prior the implementation of additional control devices and signal timing changes. The intersection approaches have two stop lines, one right at the entry where the traffic signal is located and right before crossing the intersection. The control signals are located on the first stop line across the road, and there is an additional control sign across the intersection. Also, the pedestrian crossing is located at the first stop line (Figures 3.2 and 3.3). Even though bus stops have designated areas, they are not properly used. Buses stop at any location prior, within, or after the intersection with no enforcement restrictions.

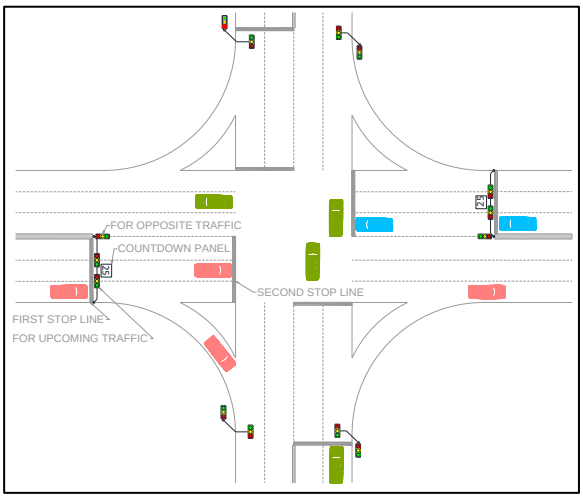


Figure 3.2. Pablo intersection scheme.

Figure 3.3. Pablo satellite image.

Some of the main concerns with the geometric design at Pablo are the lack of protection for pedestrians, the inappropriate location of signals, and lack of guidance. Left turning movements were not properly made. There was not a particular signal time interval that allowed this conflicting movement to freely move within the intersection. As a consequence, there were problems such as troublesome vehicle interactions, inappropriate encroachment and vehicles in the middle of the intersection. There appeared to be no clear understanding on how the intersection was supposed to work.

Fortuitously, countdown panels (a traffic signal device that displays the time remaining of each phase intervals) were implemented during the data collection period; thus, there was data to conduct a before and after analysis of the implementation of the device. Figure 3.4 shows the original intersection and 3.5 shows the countdown panel installed on the mast, right next to the signal head.

Figure 3.4. Pablo intersection before countdown panel implementation.

Figure 3.5. Pablo intersection after countdown panel implementation.

3.2.2. BEIJING Intersection

Beijing intersection is a signalized traffic circle. Only the east approach has a countdown panel (see Figures 3.6 and 3.8). The stop lines are marked at the entrance to the circulatory roadway. There is a designated pedestrian crossing at each stop line although there are two pedestrian bridges, one in each approach within 200 meters away from the intersection. The bus stops operate similarly to Pablo; thus, buses stop anywhere. Figure 3.6 displays the geometry and signal control devices features with their corresponding locations at Beijing intersection.

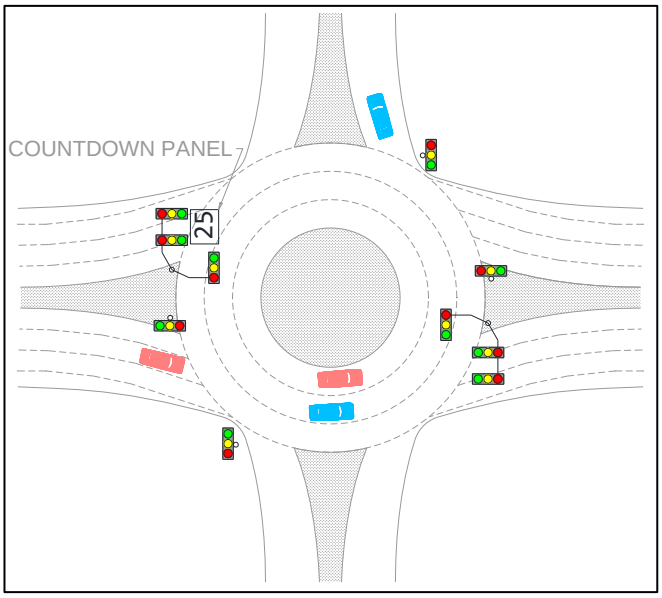


Figure 3.6. Beijing scheme.

Figure 3.7 illustrates the location of the pedestrian bridge along with the operational car distribution at the intersection. Figure 3.8 and 3.9 display the east and west approaches at Peru intersection where the locations and characteristics of the traffic signals can be observed.

Figure 3.7. Beijing satellite image.

Figure 3.8. East approach Beijing intersection (countdown panel).

Similarly, the pedestrian crossing configuration is worrisome. Since vehicle compliance to the stop line is deficient, vehicles tend to go beyond that limit closer to the conflictive traffic to depart earlier. As a consequence, pedestrians have to cross in between vehicles during the red phase. Furthermore, the long crossing distance means a

long period of time of exposure to the traffic, and vehicles from the inner lanes might not see pedestrians crossing due to occlusion.

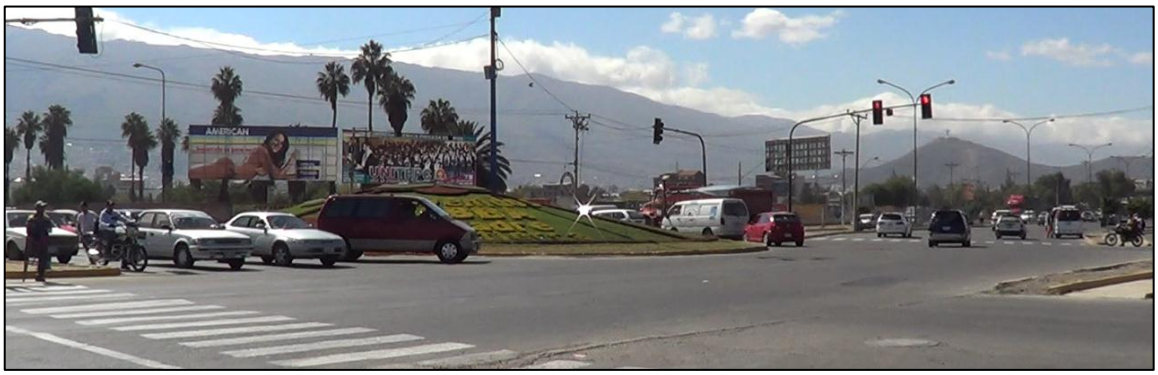


Figure 3.9. West approach Beijing intersection.

The most important characteristic of Beijing intersection is the fact that the facility is the designated intersection to accommodate heavy vehicles on their transition from the main road to the arterial and vice versa. The arterial is the designated detour for heavy and large vehicles to travel through the city. Consequently, the intersection deals with a complex set of vehicle mix which results in unstable performance and safety issues. The most identifiable problem regarding this traffic circle is the accumulation of vehicles on the circulatory roadway when they are waiting for the green light to make turning movements. Also, when large heavy vehicles enter the intersection they tend to disregard the signals and intimidate the different approaching vehicles with their size. They get in the way of major flow streams or cut across the active roadway. During data collection such complete blockage of the intersection was witnessed where no vehicle was able to move for several complete signal cycles.

3.2.3. PERU Intersection

Peru is a signalized traffic circle. It has traditional features with numerous trees in the central island, single control signal for the main approaches, and reduced right of way. The intersection has a particular characteristic: the main approaches have identifiable speed and flow differences because of the availability of space, proximity to other facilities, a more urbanized area, and parking. Similar to other intersections, there is a designated pedestrian crossing at each stop line, and there are two pedestrian bridges, one on each approach within 200 meters away from the intersection. The buses stop anywhere or at any time. Figures 3.10 illustrate the simplicity of traffic signal configuration. It contains only single masts for traffic devices for each approach to observe. In Figure 3.11 the differences of both main approaches and vegetation can be observed. Figures 3.12 and 3.13 show on site observations of the traffic signal configuration and visibility at Peru intersection.

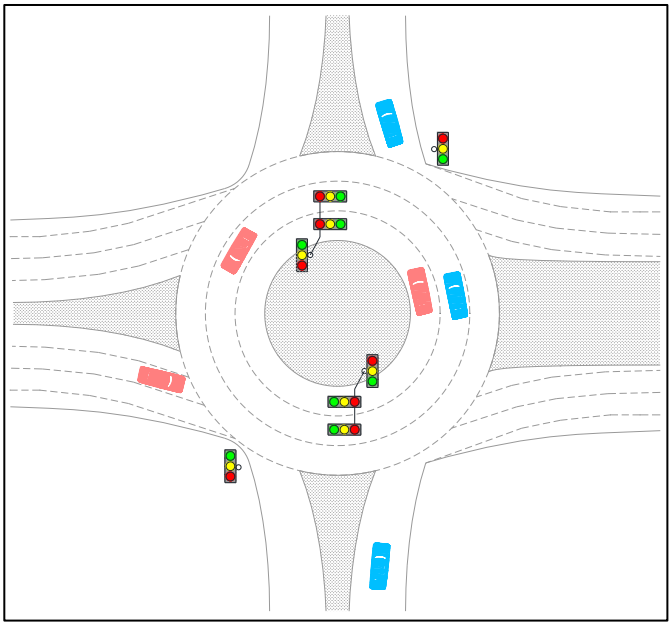


Figure 3.10. Peru scheme.

Figure 3.11. Peru satellite image.

Figure 3.12. West approach Peru intersection.

Figure 3.13. East approach Peru intersection.

3.3. Data Collection Procedure

Videos were recorded during 6-hour intervals for 7 weeks on weekdays from May to July, 2012. At each major approach, the red light running vehicles, the traffic signal, the direction of the vehicles, and the speeds were captured. The time periods of the video recordings of the traffic at the intersections were from 9:00 AM to 01:00 PM, and video recordings with speed measurements were conducted from 1:30 to 3:30 PM.

3.3.1. Intersections Traffic Video Recording

The video cameras were placed according to the traffic characteristics and the geometry of the intersections. At some intersections the traffic was filmed by facing the main roadway traffic on opposite sides of the intersection as show in Figures 3.14 and 3.15. The determination of the locations for the video recordings was primarily based on finding the best image coverage for optimal video data processing.

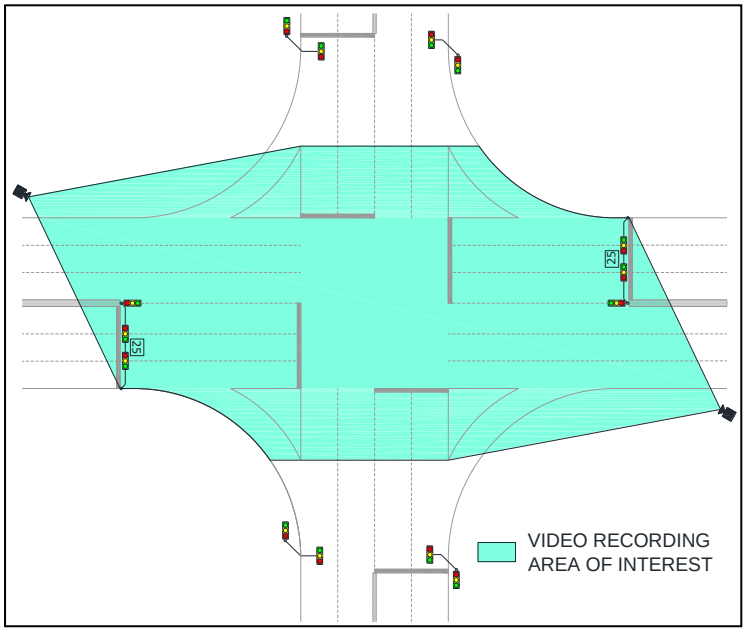


Figure 3.14. Area of interest for video recording Pablo intersection.

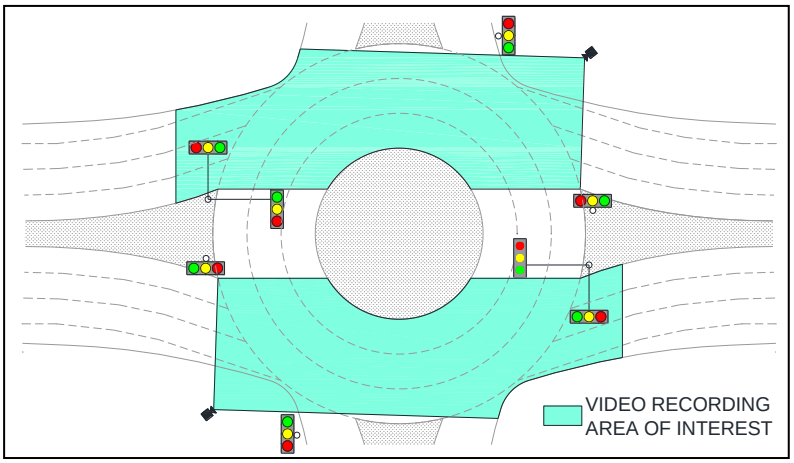


Figure 3.15. Area of interest for video recording Beijing and Peru intersection.

3.3.2. Approach Video Recording and Speed Measurement

For speed data, several locations were investigated to determine the optimal location and conditions. At first the first speed recordings were conducted from the side of the main road approach, but the radar gun captured some vehicles from the opposing road. As a result, it was difficult to identify to which stream the speed measurements belonged to. Finally, the decision to locate the video camera and radar gun on the median was used to avoid the problems observed in the previous location. As illustrated in Figures 3.16 and 3.17, the measurements were intended to record speeds of vehicles along with the video recording at 200 meters from the intersection.

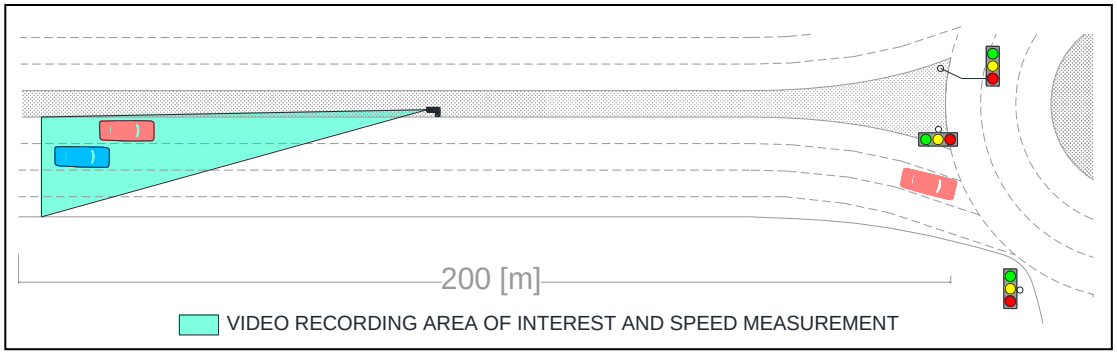


Figure 3.16. Area of interest for video recording and speed measurement.

Figure 3.17. Speed video recording procedure.

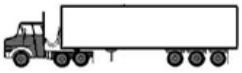
3.4. Video Processing Methodology

It is difficult to process data accurately from video when many measurable parameters are taken into account. Initially, using a 15 minute sample footage, the most reasonable and obtainable parameters were processed; however, more difficulties arose once a footage was processed. Therefore, the measurable parameters were narrowed down and simplified to the most relevant to the main purpose of the study. A brief description of the elements considered for the data processing to explain the way the data was measured, counted, and coded.

3.4.1. Traffic Flow, Red light, and Yellow Light Running Count

The count of vehicle flow was conducted according to local features of traffic, vehicles, and driving behavior. A vehicle classification was established for the purpose of an adequate identification. Table 3.1 shows a classification scheme based on four main categories.

Table 3.1. Vehicle classification.

1. Motorcycle	2. Small vehicle	3. Buses	4. Trucks
	**Passenger car, two axel, four tire vehicle**  	**(3.1) Small <15 seas**  **(3.2) Medium 15-25 seats**  **(3.3) Large >25 seats** 	**(4.1) Single unit**  **(4.2) Trailer**  

The traffic movements considered in the study are for vehicles entering the main leg and the corresponding arterial entrance. The procedure focuses on counting conflicting vehicles entering the intersection at the particular studied approach at every complete signal cycle. The other flow movements are not counted because they do not directly interact with the main approach flow. Additionally, vehicle traffic movements were classified by cases. Cases A, B, and C classify vehicle movements according to the location of entering traffic and the intersection. Case A displayed in Figure 3.18, considered vehicle movements from the main approach which are crossing through the intersection, taking a right turn, or taking a left turn in Pablo intersection.

Case B and C, observed in Figures 3.18 and 3.19, are the different vehicle movements using the intersection, but in this cases the approaches are the arterials. Case

C was not counted for Peru and Beijing intersections because it was not possible to record consistent data due to visibility issues. However, it was not necessary to record Case C because of the lack of influence on the main flow stream.

Case C does not play a determining factor for the red light running or safety on the main approach, since the geometrics of the separate right angle vehicle interaction.

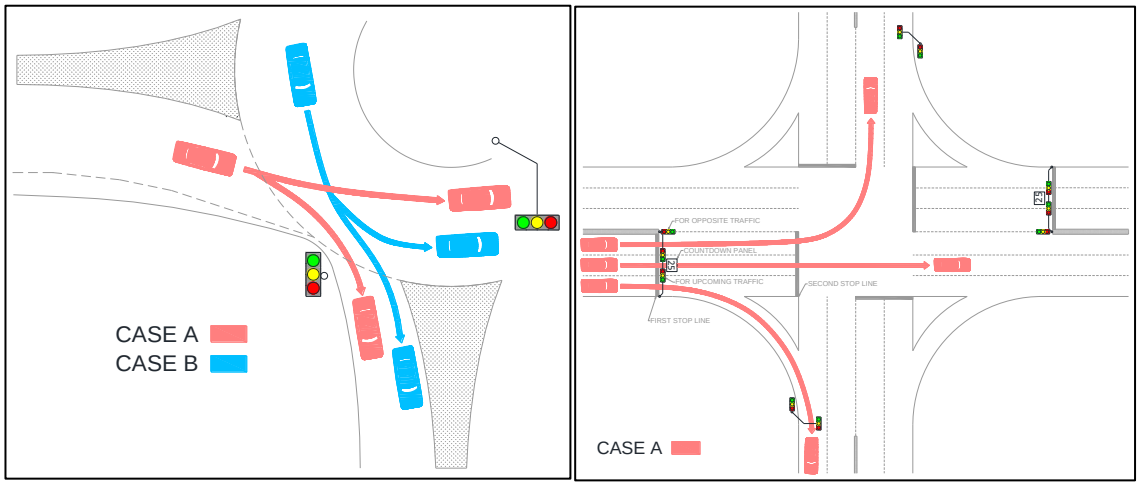


Figure 3.18. Vehicle movements.

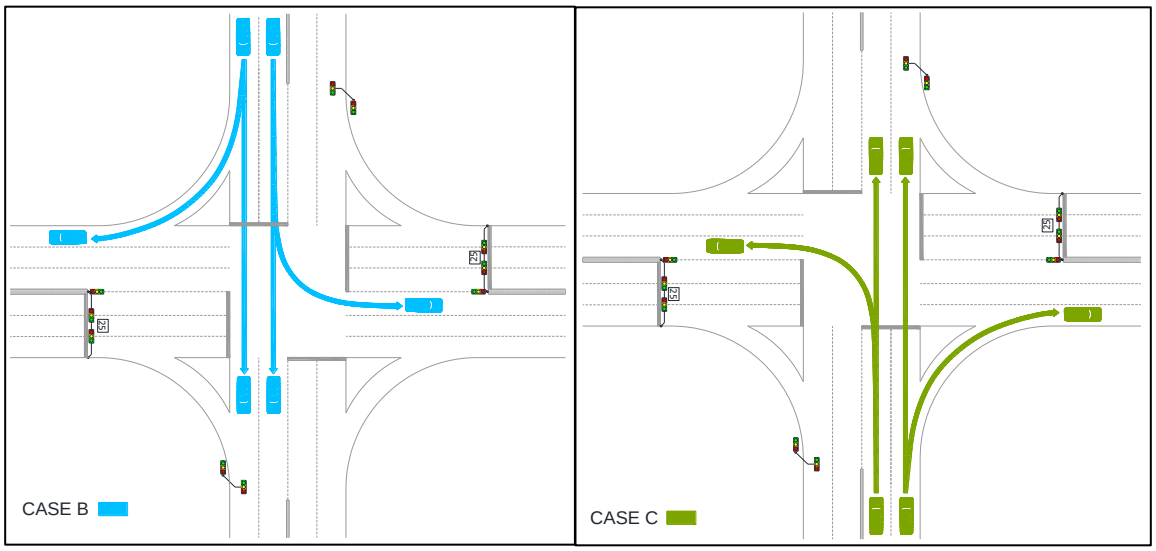


Figure 3.19. Case B and C four legged conventional intersection (Pablo intersection).

3.4.2. Approaching Speed Measurement

The speed data processing was conducted following lane and vehicle classification using video recordings. It was determined that two hours would cover the minimum sample size for most vehicle groups. However, there is one vehicle group that is not very common, so it was not included in individual group speed analysis (Group 3.3 large bus). Only one or two measurements were recorded of that vehicle group per hour.

3.4.3. Procedure and Measurement Storage

The flow, signal violations, and speed measurements were stored for each complete signal cycle according to the vehicle classification, turning movement, or lane location, in the case of the speed data. For the flow and signal violations count, a spreadsheet observed in Figure 3.20 was generated by using single keys from the keyboard. Figure 3.21 shows the distribution of the keyboard according to vehicle groups and movements. The count was stored in each cell as a continuous set of text that was later counted and classified in different cells using a counting function. As a result, a specific count of the different variables of interest was obtained for every cycle of video processing conducted.

5/31/2012								
CYCLE	RECORDING TIME	INTERVAL CONTROL	VOLUME COUNT		YELLOW RUNNING COUNT		RED LIGHT RUNNING COUNT	
			A	B	A	B	A	B
1	0:00:22	0:00:00	AADAAAAZAA	AGASGWWQ(	AJ	G		Q
2	0:01:29	0:01:07	AAAAAAAZZA(	ASASSAAYQW	AAA			
3	0:02:36	0:01:07	AAGAAAAASA	AAQQWQAAQ	A	QQ		
4	0:03:43	0:01:07	DDAASZZZAZ[	QAWQQQAW	AA	AAS	Z	QQ
5	0:04:50	0:01:07	AAAAADZCAA	AAQQAAAAAC	AA	T	A	QAT
6	0:05:57	0:01:07	GAAASAAAZZ	AQWQWDAW	AAAAAWWQA	AA	Z	

Figure 3.20. Volume and signal violation count spread sheet.

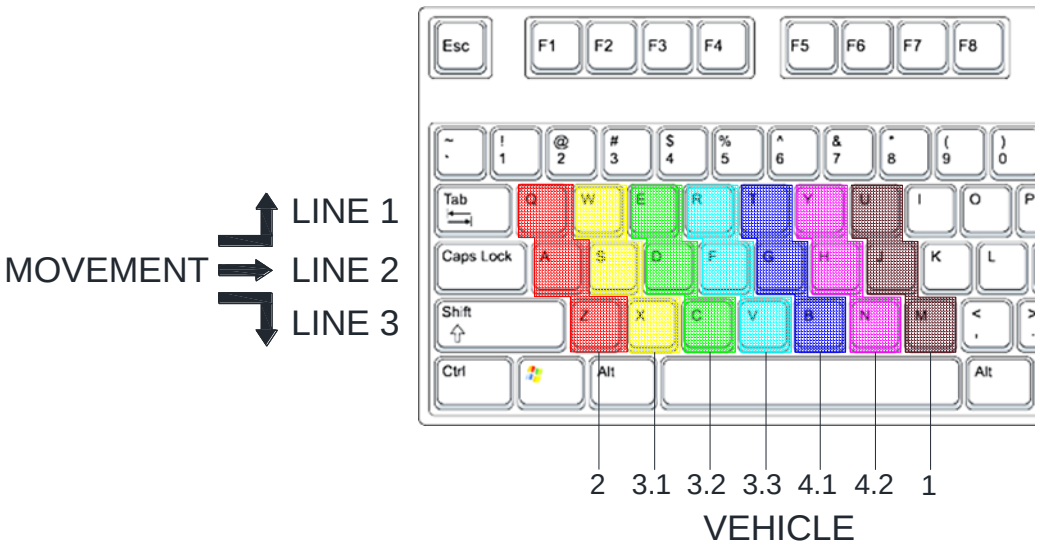


Figure 3.21. Keyboard configuration for volume and signal violation count.

Similarly for the speed data storage, the vehicle classification and lane were entered into a cell as a single set of numbers as illustrated in Figure 3.22. Later the set of numbers were classified in different cells to provide each variable for each observation. In this case, the vehicle classification was coded from numbers 1 to 7 according to the groups already determined in previous count procedure. The lanes were coded from 1 to 3: 1=inner, 2=middle, and 3=outer lane. The format was the lane as the first digit, then the vehicle as the second digit, and finally the rest of the digits represented the speed recorded.

		7/17/2012			
INPUT	LANE	VEHICLE	SPEED	REC TIME	TIME
1762	1	7	62	0:00:00	1:00 PM
2155	2	1	55		
3138	3	1	38		
1162	1	1	62		

Figure 3.22. Spread sheet configuration for speed measurement.

3.5. Summary of Data Processing

The database for the analysis of the red light running at traffic circles included volume counts, yellow and red light running counts, and speed measurements. During the summer of 2012, approximately 110 hours were recorded at different locations by following the methodology described previously.

An overall summary of the intersection parameters is presented in Table 3.2, including the different signal control characteristics and the length of the cycles. It is observed from that the all yellow intervals are the same at all intersections, and in general the cycles length according to the significance of traffic flow at each location. The distance from each intersection to neighboring intersections along the corridor is described in Table 3.3. The distance to close facilities was considered for analysis of the approaching speed.

Table 3.2. Summary of signal cycle timing.

Int.	App.	Signal	Cycle length (s)	Green interval (s)	Yellow interval (s)	Red interval (s)
Pablo	West	Traffic light	75	45	3	27
		Traffic light and panel				
	East	Traffic light	90	57	3	30
		Traffic light and panel	75	45	3	27
Beijing	West	Traffic light	76	45	3	28
	East	Traffic light and panel				
Peru	West	Traffic light	67	39	3	25

Table 3.3. Summary of distances to other intersections.

Int.	App.	Additional roadway	Approach length (m)	To intersection	Lanes	Speed limit (km/h)
Pablo	West	Shoulder	900	Sexta	3	70
	East	Shoulder	1600	Beijing		
Beijing	West	Shoulder	1600	Pablo		
	East	Shoulder	1500	Peru		
Peru	West	Shoulder	1500	Beijing		
	East	Parking lane	350	Puente		

After the initial learning curve, a processing time of around 8 hours per approach was reached in order to make a proper analysis. From the video recordings, the ones with the best qualities were selected for data processing. A summary of the hours of video processing, vehicles counted, yellow light running, and red light running, is shown in Table 3.4. It is observed that over 70 hours were expended in processing for volume. Table 3.5 shows a significant number of yellow and red-light running observations at each intersection and approach.

Table 3.4. Summary of volume video processing and observations.

Int.	App.	Signal	Video volume processing (hr)	Observations	
				Vehicles	Cycles
Pablo	West	Traffic light	7.6	20639	370
		Traffic light and panel	8	20734	382
	East	Traffic light	7	17554	280
		Traffic light and panel	8	18391	382
Beijing	West	Traffic light	12	33880	570
	East	Traffic light and panel	12	32442	570
Peru	West	Traffic light	8	20809	432
	East	Traffic light	8	20860	432

Table 3.5. Summary yellow and red light running observations.

Int.	App.	Signal	Observations							
			Yellow-light-run				Red-light-run			
			Case A	Case B	Case C	Case AA	Case A	Case B	Case C	Case AA
Pablo	West	TL	427	214	150	519	650	116	181	456
		TLP	350	223	143	499	562	116	164	323
	East	TL	356	146	143	310	716	125	79	368
		TLP	300	223	267	391	841	161	360	288
Beijing	West	TL	569	622			167	268		
	East	TLP	502	782			126	417		
Peru	West	TL	706	660			218	367		
	East	TL	665	676			162	483		

TL: Traffic light, TLP: Traffic light and panel

Similarly, for speed measurements, 2 hours of video recording produced a large enough sample for the speeds. Tables 3.6 contains a summary of the total hours of video processing for speed measurements, and over 20 hours were expended in processing for speeds.

Table 3.6. Summary yellow and red light running observations.

Int.	App.	Signal	Video speed processing (hr)	Observations
				Speeds
Pablo	West	Traffic light	2.00	4451
		Traffic light and panel	2.00	3465
	East	Traffic light	3.25	5285
		Traffic light and panel	2.00	2929
Beijing	West	Traffic light	4.00	7793
	East	Traffic light and panel	4.25	8165
Peru	West	Traffic light	2.00	3518
	East	Traffic light	2.00	2991

CHAPTER 4. DATA ANALISYS

4.1. Introduction

The analysis of the data covers an overall descriptive statistics analysis, comparisons of different parameters using hypothesis testing, and the development of a model to predict red light running rates based on variables obtained from the data collection.

4.2. Descriptive Statistics

This section focuses on determining predominant conditions and patterns regarding volume, vehicle mix, yellow light running, and red light running.

4.2.1. Traffic Volume

The volume found from the dataset is based on the total number of vehicles observed during the period of study at each location. The traffic volumes distributed in their corresponding movements allows the visualization of the predominant direction of travel and the functionality of each intersection As shown in Table 4.1, the major approaches have average volumes between 1600 and 1900 vehicles per hour. On the other hand, the minor approaches generate small volumes which interact with the main road traffic. In Figures 4.2 and 4.3 the left turning movement ranges from 250 to 500 vehicles per hour at the traffic circles. However, in Figure 4.1, at the four legged intersection, the left turning movements are not significant. They range from 20 to 70 vehicles per hour.

Table 4.1. Average entering volume.

Int.	App.	Average volume (veh/hr)								
		Case A			Case B			Case C		
		Left	Through	Right	Left	Through	Right	Left	Through	Right
Pablo	East	28	1496	132	39	311	60	34	283	230
		1655			410			547		
	West	73	1704	155	34	283	230	39	311	60
		1933			547			410		
Beijing	East		1569	227	246	654				
			1797		900					
	West		1803	106	294	613				
			1909		907					
Peru	East		1404	226	496	469				
			1629		965					
	West		1518	196	387	487				
			1714		874					

Table 4.2. Average exiting volume.

Intersection	Approach	Exit to	Average volume (veh/hr)
Pablo	East	North	516
		West	1765
		South	466
	West	North	516
		East	1798
		South	466
Beijing	East	North	695
		West	1900
	West	South	683
		East	1905
Peru	East	North	881
		West	1815
	West	South	719
		East	2097

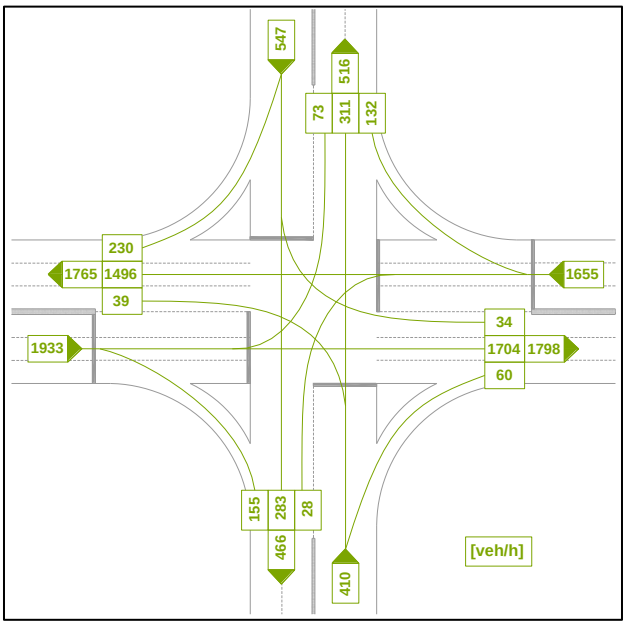


Figure 4.1. Average volume (Pablo intersection).

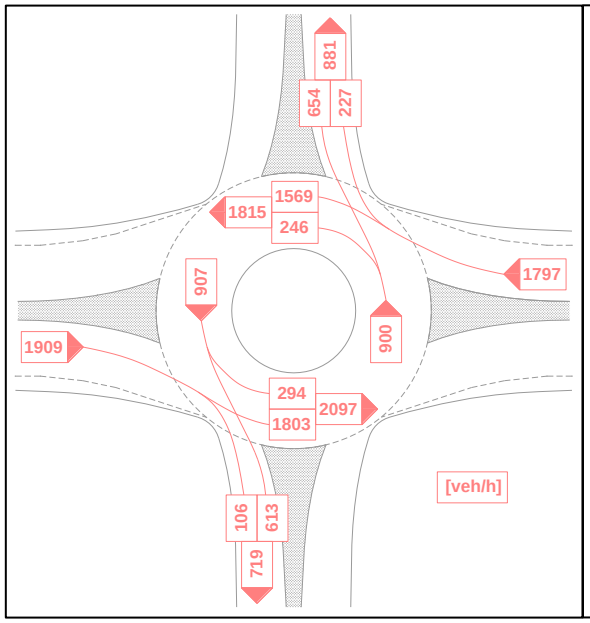


Figure 4.2. Average volume (Beijing).

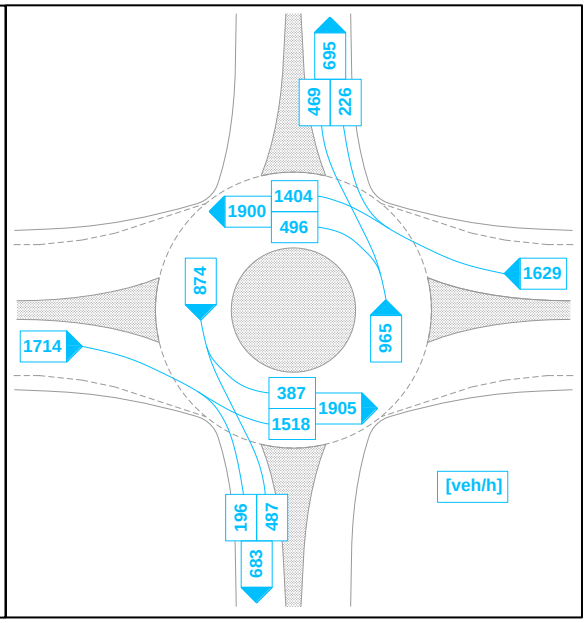


Figure 4.3. Average volume (Peru).

Beijing intersection deals with highly distributed traffic movements where there is considerable accumulation of queue on the circulatory roadway due to high volumes of turning movements stopped by the signal timing. Similarly, Peru intersection has

considerable traffic volume on the circulatory roadway that either diverges or turns around the intersection.

4.2.1.1. Vehicle Mix

For the main road distribution, the predominant circulating vehicle is the small vehicle (62.81%) and buses represent the second largest vehicle group (18.07%). Bus driver behavior is considered the worst by the local community. They are aggressive and reckless, and violate traffic rules the most. Table 4.3, shows the percentages of vehicle mix distribution.

Table 4.3. Vehicle classification.

Classification	Group	Subgroup	%	Type of vehicle
1	Motorcycle	1	5.77	Motorcycle
2	Vehicle	2	62.81	Small vehicle
3	Bus	3.1	18.07	Small size bus
		3.2	4.86	Medium size bus
		3.3	0.28	Large size bus
4	Truck	4.1	7.16	Single unit truck
		4.2	1.05	Multiunit trailer

Figures 4.4 and 4.5 contain a graphical classification display of the vehicle groups of buses and trucks to effectively visualize its presence in the roadway system that travels across the intersections. As mentioned before, there is a significant presence of buses and heavy vehicles that circulate in the corridor.

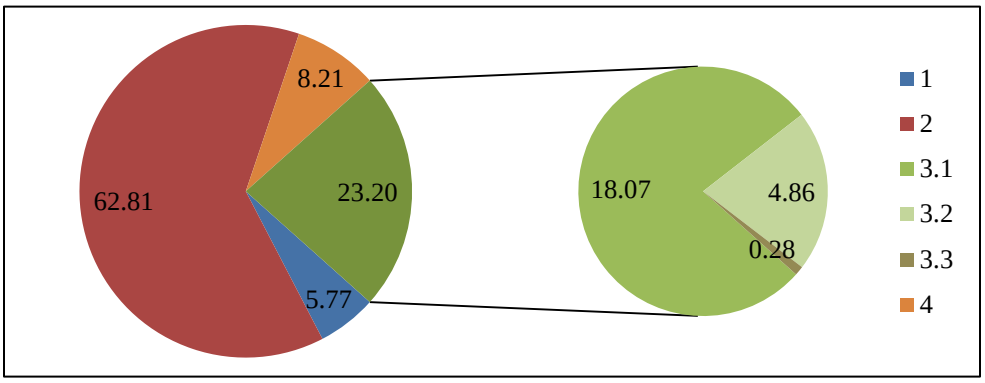


Figure 4.4. Main road bus distribution.

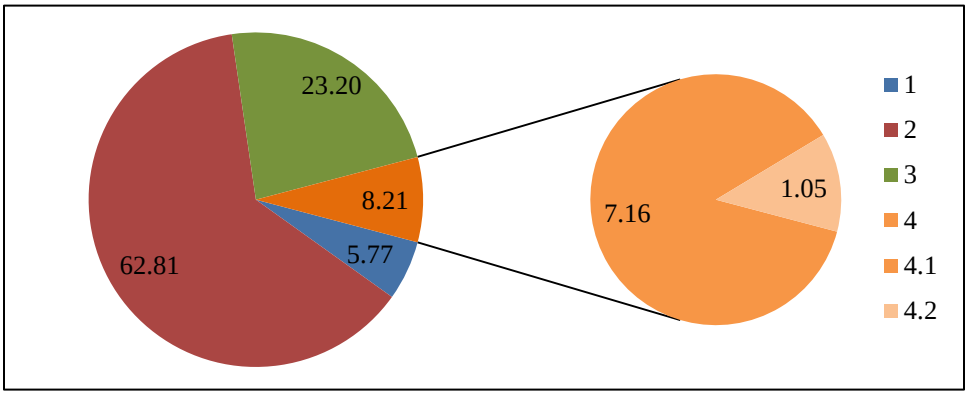


Figure 4.5. Main road truck distribution.

Vehicle mix according to the intersections has a different pattern because of the geometry of the intersections, the rules applied to heavy vehicles, and the routes of transit for their circulation along the main road. For instance, Pablo and Beijing intersections have the largest presence of small public buses with 21.27 % and 18.85% of the total traffic in Case A observed in Table 4.4. These two intersections are mainly used by these public buses to turn north or south from the main road to connect to arterials before traveling further into the congested downtown area. Also, there is a very small percentage of Group 3.2 at Beijing intersection, only 0.38% in Case B as observed in Table 4.4. The main reason for this is that medium size buses are predominantly used to

circulate directly to the downtown area without any detour or using alternative arterials. Regarding Peru intersection, the presence of large vehicles like Groups 3.3, 4.1, and 4.2 is small because this intersection is close to more populated areas with less roadway space and limited altitude clearance. Finally, the presence of heavy vehicles in both Pablo and Beijing ranges between 6.89% to 9.37%, as observed in Table 4.4. Heavy cargo transportation is limited until Beijing intersection where the detour starts for some heavy trucks and all type of trailers.

Table 4.4. Intersections vehicle subgroup distribution.

Intersection	**Case**	**Percentage of vehicle subgroups (%)**						
		1	**2**	**3.1**	**3.2**	**3.3**	**4.1**	**4.2**
Pablo	Case A	5.05	58.02	21.27	4.64	0.34	9.37	1.30
	Case B and C	9.87	74.71	5.99	1.36	0.04	7.75	0.27
Beijing	Case A	5.96	62.45	18.85	4.25	0.35	6.89	1.24
	Case B	8.78	72.84	6.86	0.38	0.46	7.87	2.82
Peru	Case A	6.95	73.23	10.21	6.30	0.02	3.09	0.20
	Case B	8.71	68.85	16.22	2.26	0.02	3.55	0.40

4.2.1.2. Yellow and Red Light Running Rates

The light running rates are linked to cycle, hours, and total vehicles as listed in Table 4.5. The units used are: light running per hour (LR per hr), light running per 100 signal cycles (LR per 100 cycles), and light running per 1000 vehicles (LR per 1000 veh). For this research, the units light runners per hour were chosen because it relates to traffic flow data variables collected.

Table 4.5. Intersections vehicle subgroup distribution.

Approach	YLR	RLR	Cycle	hr	Veh.	YLR per hr	RLR per hr	YLR per 100 cycles	RLR per 100 cycles	YLR per 1000 veh.	RLR per 1000 veh.
Case A											
Pablo EC	391	288	384	8.0	13090	48.9	36.0	101.8	75.0	29.9	22.0
Pablo ENC	313	368	280	7.0	11734	44.7	52.6	111.8	131.4	26.7	31.4
Pablo WC	499	323	382	8.0	15424	62.4	40.4	130.6	84.6	32.4	20.9
Pablo WNC	519	456	370	7.7	14857	67.4	59.2	140.3	123.2	34.9	30.7
Beijing E	523	128	570	12.0	22088	43.6	10.7	91.8	22.5	23.7	5.8
Beijing W	587	181	570	12.0	23453	48.9	15.1	103.0	31.8	25.0	7.7
Peru E	665	162	432	8.0	13101	83.1	20.3	153.9	37.5	50.8	12.4
Peru W	706	218	432	8.0	13783	88.3	27.3	163.4	50.5	51.2	15.8
Case B											
Pablo E	1101	1186	664	30.0	13456	36.7	39.5	165.8	178.6	81.8	88.1
Pablo W	983	1006	752	31.4	14829	31.3	32.0	130.7	133.8	66.3	67.8
Beijing E	802	435	570	12.0	11048	66.8	36.3	140.7	76.3	72.6	39.4
Beijing W	638	283	570	12.0	11145	53.2	23.6	111.9	49.6	57.2	25.4
Peru E	676	483	432	8.0	7759	84.5	60.4	156.5	111.8	87.1	62.3
Peru W	660	367	432	8.0	7026	82.5	45.9	152.8	85.0	93.9	52.2

E: East, W: West, C: Counter, NC: No counter YLR: Yellow light running, RLR: Red light running

The identifiable patterns observed from the results in Table 4.5 are:

- There are slightly higher yellow and red violations rates coming from the west approaches (Case A).
- There are higher light violations with no counter down panels meaning that there is possibly a significant improvement on light violations due to their implementation.
- The violation rates for all the approaches regarding case B are considerable larger than the case A. In some cases two to four times larger.

- Most arterials directly linked to the east approach present larger violation rates in both yellow and red lights (coming from the south or left turns from the main roadway).

4.2.2. Speed Data

One of the main interests in collecting approach speeds is to analyze any significant variables to red light running, and changes due to the different scenarios, circumstances, or control devices in place. The different descriptive statistics of the speed measurements is shown regarding the approach, lane, and vehicle distribution.

4.2.2.1. Speed by Approach

Approaching speed statistics have been determined for each of the main approaches to the three intersections under study. The differences in operations, control systems, distance, geometry, and surroundings of the different approaches are analyzed to determine any contribution to a particular speed driving behavior. Table 4.6 shows the following:

- Pablo intersection has higher approaching speeds at the east approach.
- At Pablo intersection, there is little difference with the speed measurements with and without the countdown panels.
- The highest mean speeds recorded are at Pablo intersection.
- At Beijing intersection, the east approach has a countdown panel, and there is no apparent difference in the approaching speed with the west approach that does not have the countdown panel.

- Peru intersection has considerably higher mean speed from the west approach compared to the east approach. This can be caused by the parking lane, wider median, short distance to a close by facility, and a more urbanized region at the east of the intersection.
- The east approach at Peru intersection has the lowest mean speed recorded.

Table 4.6. Statistical calculations of speed by approach (km/hr).

Approach	N	Mean	85th Perc.	Std. Dev.	Min	Max	Skewness	Kurtosis
Pablo EC	2929	63.07	76	12.90	25	156	0.27	0.98
Pablo ENC	5282	63.58	76	12.26	27	126	0.08	0.28
Pablo WC	3464	57.99	71	11.70	24	104	0.14	-0.52
Pablo WNC	4451	58.66	69	10.62	25	149	0.21	1.06
Beijing EC	6559	61.49	73	11.83	24	118	0.08	0.13
Beijing W	5006	60.75	72	11.64	26	110	0.12	0.27
Peru E	2989	46.55	54	7.80	23	74	0.02	0.10
Peru W	3517	61.48	73	11.55	29	115	0.06	0.33

RLR: Red Light Running, C: Counter, NC: No Counter, E: East, W: West

Overall, the mean approach speeds at the different intersections do not fluctuate considerably due to geometry or signal control differences. However, there is one specific mean speed difference because of particular features of the approach observed east of Peru intersection when only a maximum speed value of 74 (km/hr) was recorded. The presence of a wider median, parking lane, and a close bridge facility are characteristics that set the conditions for lower approaching speed.

4.2.2.2. Speed by Lane

The speeds of vehicles at different lanes on the main roadway were determined. The results follow an expected and common speed lane distribution where the higher

speeds occur on the inner lane and decrease towards the outer lane. Although the mean speed resemble such trend, it can be observed in Table 4.7 that the maximum speed values can occur on any lane.

Table 4.7. Statistical calculations of speed by lane.

Lane	N	Mean	85th Percentile	Std. Dev.	Min	Max	Skewness	Kurtosis
Inner	13644	65.4	76	11.00	28	156	0.20	0.83
Medium	11891	59.4	71	11.52	24	113	0.19	0.06
Outer	8662	51.7	63	10.53	23	149	0.43	1.05

4.2.2.3. Speed by Type of Vehicle

As part of the classification of the vehicles, speed has been considered as criteria to judge the relevance of the vehicle according to their characteristics, use, and operability. Therefore, small vehicles (Group 2) have the highest mean speed, and it goes down according to the size increment and shape until lower mean speed values with large size vehicles (Groups 3.2 and 4.2). On the other hand, it is a concern that the maximum values recorded are for vehicle groups 3.1 and 3.2, which are public service buses as observed in Table 4.8.

Table 4.8. Statistical calculations of speed by vehicle (km/hr).

Veh.	N	Mean	Median	85th Perc.	Std. Dev.	Min	Max	Skewness	Kurtosis
1	1868	57.6	57	70	12.25	25	118	0.40	0.22
2	22172	62.7	63	75	11.76	24	126	0.11	0.21
3.1	5888	56.0	56	67	10.52	25	156	0.37	1.27
3.2	1412	46.2	45	58	10.71	23	149	1.09	5.60
3.3	123	57.3	58	68	10.61	29	85	-0.34	0.16
4.1	2363	53.7	53	65	11.17	25	100	0.44	0.24
4.2	371	49.3	49	60	10.02	24	81	0.21	0.01

The speed data follows a normal distribution and the variability parameters are fairly constant as it can be verified with parameters of central tendency and variability displayed in Table 4.8. Little difference has been observed regarding approach direction and countdown panel presence. Also, no major differences have been found between intersections. However, the east approach of Peru intersection presents particular circumstances that considerably influence the approaching speed. As expected, the lane speed distribution varies in descending order from the inner to the outer lane. Finally, it is a concern that the maximum values recorded for public buses were as high as the ones of a small vehicle. In the following section, a complete analysis using hypothesis testing to determine significant differences between the observed parameters was conducted to support the statements made regarding possible differences.

4.3. Comparisons of Parameters

In the previous section, different parameters were presented in order to identify potential differences and patterns. The primary sources of comparisons were based on speed differential, yellow running rate, control devices, and geometric countermeasures. Hypothesis testing was used to determined significant differences in both red light running and speed. The objective of finding significant differences is to determine if the local administration policies leads to efficient performance and safety, and if the countermeasures implemented contribute to the overall system of intersections in the segment under analysis.

Regarding the hypothesis testing methods, the two sample z-test has been used for the speed data because the sample sizes range from 2929 to 6569 observations, and

the two sample t-test were used for the red light running rates (RLR) because the sample sizes range from 28 to 48 observations. The units used in the comparisons are: Speed (km/hr), Red Light Running (RLR/hr), and Yellow Light Running (YLR/hr).

The most demanding work was to aggregate the red light running data, which was recorded in vehicles per signal cycle to vehicles per hour with the different signal timings. The data was initially grouped for every 15-minute time interval, and the calculation of traffic flow was performed in units of vehicles per hour. A summary of the aggregated data is presented in Appendix A.

4.3.1. Comparisons of Red Light Running Based on Significant Speed Difference

Intersection Peru (traffic circle) is used to analyze red light running trend regarding significant difference on the approaching speed. The east approach has an average of 46.55 (km/hr) with a standard deviation of 7.80 (km/hr). The west approach has an average of 61.48 (km/hr) with a standard deviation of 11.55 (km/hr). From both averages, there is an important speed differential of 14.93 (km/hr) (highest speed differential from all intersections under study).

Initially, the significance of the speed difference is conducted. Considering the normal distribution and large speed dataset, the z-test was used in this particular case as illustrated in Table 4.9. There is significant statistical evidence that the average approach speed from the west is larger than the east approach. The influence of the geometrics and distances to nearby intersections are considered to be the main factors contributing to the difference in the speed from the two main approaches to this intersection. The east

approach does not have a long distance to the next facility, it has a parking lane, and its traffic flow is relatively low. Therefore, these features limit higher approaching speeds.

Table 4.9. Hypothesis test for speed differential.

Test	Comparison	Int.	App.	H_O	H_A	P-value
Two sample z-test	Speed differential	Peru	E	Speed W = Speed E	Speed W > Speed E	<0.0001
			W			

E: East, W: West

The red light running rates calculated at this location are also tested as shown in Table 4.10. The comparison test the significance of red light running at the east approach with speeds [60.40 (RLR/hr)] being greater than the west approach with lower speeds [45.90 (RLR/hr)]. The evidence of the hypothesis test suggests that red light running was higher at the west approach where higher speeds are recorded.

Table 4.10. Hypothesis test for red light running with speed differential.

Test	Comparison	Int.	App.	H_O	H_A	P-value
Two sample t-test	RLR with speed differential	Peru	E	RLR-HS = RLR-LS	RLR-HS > RLR-LS	0.0010
			W			

RLR: Red Light Running, HS: Higher Speed, LS: Lower Speed

4.3.2. Comparison of Red Light Running Based on Traffic Control Signal Differences

4.3.2.1. Countdown Panel Usage at Traffic Circles (Long Term)

At Beijing intersection (traffic circle), the east approach has a countdown panel combined with a traffic signal, but the west approach only has a traffic signal. The signal configuration has been in place for more than five months, from January 2012 until the collection of the data in June of 2012. The values to be tested were: mean (standard

deviation), east approach 14.91 (15.38) (RLR/hr) and west approach 10.86 (8.82) (RLR/hr), and the result is shown in Table 4.11. There is no statistical evidence that there is difference in red light running rates at traffic circles with the usage of countdown panels on a long-term basis.

Table 4.11. Hypothesis test for countdown panel usage at traffic circles.

Test	Comparison	Int.	App.	H_O	H_A	P-value
Two sample t-test	Count down panel usage	Beijing	E	RLR-C = RLR-NC	RLR-C ≠ RLR-NC	0.1201
			W			

RLR: Red Light Running, C: Counter, NC: No Counter, E: East, W: West

4.3.2.2. Before and After Countdown Panel at Four Leg Intersection (Short Term)

Pablo intersection has gone through geometric changes and control devices modifications from a convectional traffic circle. Currently, it is a four-legged signalized intersection. When the data was collected at the location in early June of 2012, the presence of countdown panels in the control signal system were not in place. Around 6 weeks later, in late July 2012, when the intersection was revisited for site verification purposes, the countdown panel devices had been in operation for 4 weeks. Therefore, due to such fortunate timing, a before and after comparison of the implementation was possible.

The values tested were: mean (standard deviation), east with no counter 52.57 (18.56) (RLR/hr), east with counter 36.00 (14.69) (RLR/hr), west with no counter 58.87 (18.10) (RLR/hr), and west with counter 40.38 (14.08) (RLR/hr). With the implementation of the countdown panels in both approaches, there was a significant reduction in red light running violations on a short-term basis, as shown in Table 4.12.

Table 4.12. Hypothesis test for before and after countdown panel implementation.

Test	Comparison	Int.	App.	H_O	H_A	P-value
Two sample t-test	Before and after panel implementation	Pablo	E	RLR-BC= RLR-AC	RLR-BC > RLR-AC	0.0002
			W	RLR-BC= RLR-AC	RLR-BC > RLR-AC	<0.0001

RLR: Red Light Running, B: Before, A: After, C: Counter, E: East, W: West

4.3.3. Comparison of Speeds at the Different Intersections

The purpose of the analysis of speed focuses on their variability according to particular features at the different approaches and implementations. As was mentioned previously, the two-sample z-test was used for the hypothesis testing of the speed measurements, as displayed in Tables 4.13 and 4.14. Since the sample sizes were large and a little complex to manipulate for calculations, the statistical software SAS was used to make the required calculations. Appendix A. contains the output and code of the software.

Pablo intersection approaching speeds were tested, as observed in Table 4.14. The difference in speeds between both approaches is significant with either countdown panel presence. Also, the before and after comparison of counter down panel implementation was significant in the west approach. Beijing intersection was used to test the significance of the countdown panel in one of its approaches. The result was that there was no evidence that the speeds at the traffic circle have a significant variation with the presence of a countdown panel. Finally, Peru intersection speed differential from the two main approaches was tested, and the result was discussed in section 4.2.

Table 4.13. Speed values tested (km/h).

Approach	N	Mean	Standard deviation
Pablo EC	2929	63.07	12.90
Pablo ENC	5282	63.58	12.26
Pablo WC	3464	57.99	11.70
Pablo WNC	4451	58.66	10.62
Beijing EC	6559	61.49	11.83
Beijing W	5006	60.75	11.64
Peru E	2989	46.55	7.80
Peru W	3517	61.48	11.55

Table 4.14. Hypothesis test for of speed before and after countdown panel.

Test	Comparison	Int.	App.	H_o	H_A	P-value
Two sample z-test	Speed differential of approaches	Pablo	E	S-ENC = S-WNC	S-ENC ≠ S-WNC	< 0.0001
			W			
Two sample z-test	Before and after panel speed differential	Pablo	E	S-ENC = S-EC	S-ENC ≠ S-EC	0.0811
			W	S-WNC = S-WC	S-WNC > S-WC	0.0044
Two sample z-test	Speed differential w/without panel	Beijing	E	S-EC = S-WNC	S-EC ≠ S-WNC	0.0008
			W			
Two sample z-test	Speed differential	Peru	E	S-E = S-W	S-E > S-W	< 0.0001
			W			

S: Speed, C: Counter, NC: No Counter, E: East, W: West

4.4. Model Development

4.4.1. Introduction

There is evidence to suggest that higher volumes of traffic proportionally increase of the frequency of red light violations. A multivariable linear regression model form was selected because of the relationship of the dependent variable with the independent variables which are expressed in occurrences per hour. Also, the objective

was to determine the most significant independent variables that contribute the most and are the most relevant for estimating the occurrence of red light running.

A total of 71 hours of traffic flow with vehicle classification at the three specific locations was the dataset for the model. A summary of the data can be observed in Appendix A. Initially, the data was classified by type of vehicle and number of red violations per cycle of signal, and then the data was grouped into vehicles per hour for the following analysis. The different independent variables were considered for the model, and the variables that were more significant are in the final model. Also, the condition in which the analysis was conducted was for the main approach only (Case A).

The variables considered for the analysis are the following:

R Red light runner per hour,

Y Yellow light runner per hour,

A Small vehicles per hour,

S Small buses per hour,

D Medium size buses per hour,

F Large buses per hour,

G Small single unit trucks per hour,

H Trailers per hour,

V Total vehicles per hour,

W,Z Dummy variables.

4.4.2. Model Selection Procedures

Multiple methods were used to proceed with the development and selection of the regression model. Initially, in Figure 4.6 a scatterplot of the base variables was analyzed as a preliminary observation of the behavior of the data and relationships of the variables with each other. In order to determine what variables were the most significant to be included in the model, a model selection procedure was used. The methods are all possible models, stepwise, forward, and backward selection.

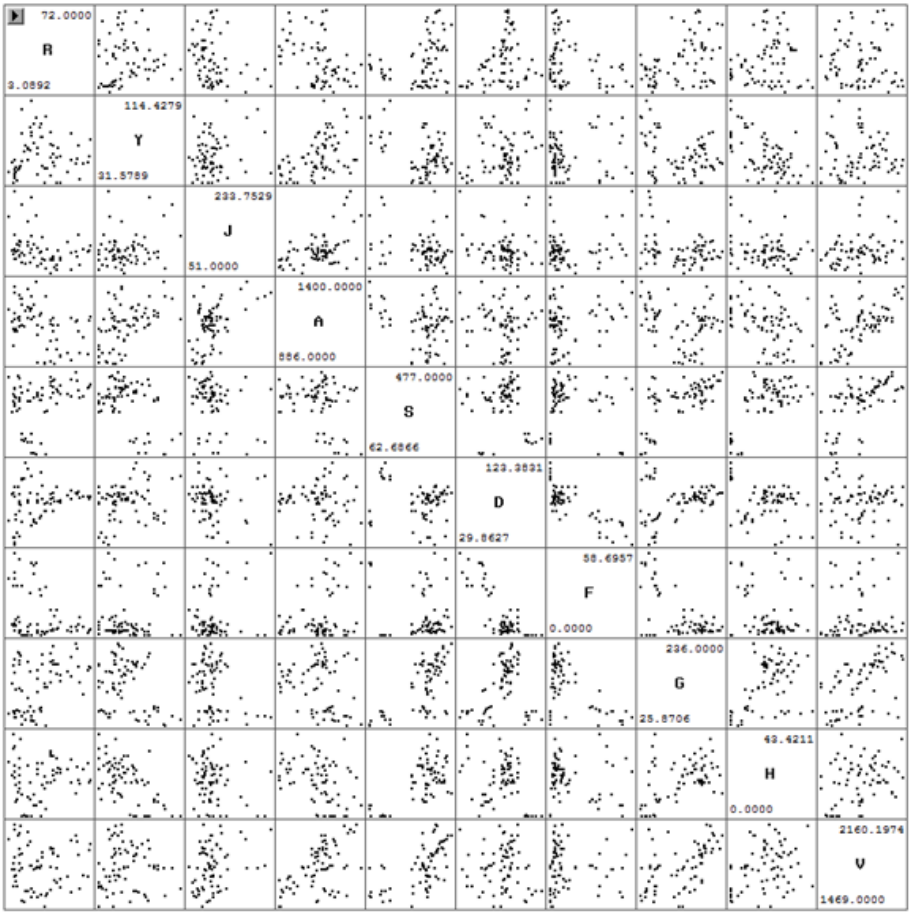


Figure 4.6. Scatterplot of all variables.

4.4.2.1. All Possible Regressions Procedure

This method guarantees the discovery of the model having the largest R^2, the smallest Cp, and so on. The procedure follows a numerical optimization to find the best model for the particular sample. However, using this method does not guarantee finding the correct model because the inclusion of more variables could result in larger measures of fit parameters that are more complex and difficult to deal with (Kleinbaum et al. 2008). Many candidates for the best model were found and considered for the selection criteria. The first 20 models were selected from the final result, and they are listed on Table 4.15.

Table 4.15. All possible regressions procedure results.

Number in model	Adjusted R^2	R^2	C_p	Variables in model
3	0.446	0.470	2.857	Y A G
5	0.446	0.486	4.915	Y A D F G
6	0.446	0.494	5.932	Y A F G H V
5	0.446	0.486	4.916	Y A G H V
4	0.446	0.478	3.922	Y A F G
7	0.446	0.501	7.007	Y J A S F H V
4	0.446	0.477	3.968	Y A G V
5	0.445	0.485	5.056	Y A F G H
6	0.444	0.492	6.171	Y J A S H V
7	0.444	0.499	7.200	Y A D F G H V
6	0.443	0.491	6.239	Y J A D F G
6	0.443	0.491	6.281	Y A D F G V
7	0.443	0.499	7.300	Y J S D F G V
4	0.443	0.475	4.294	Y A S G
6	0.442	0.490	6.429	Y A D F G H
6	0.441	0.489	6.509	Y J A F G H
7	0.441	0.497	7.524	Y A S F G H V
6	0.441	0.489	6.518	Y A S D F G
5	0.441	0.481	5.514	Y A F G V
5	0.441	0.481	5.523	Y J A F G

4.4.2.2. Backward Elimination Procedure

The procedure selection initially contains all variables in the model, and it determines the partial F tests. The selection focuses on the lowest partial F statistic. So according to the significance level specified, it was decided whether to remove the variable from the model under consideration. The procedure was stopped once there were no more non-significant variables to be removed (Kleinbaum et al. 2008). For this procedure a confidence level of α=0.05 was used. In Table 4.16, the procedure was summarized. The remaining variables as a result of the procedure are: A, Y, and G.

Table 4.16. Summary of backward elimination.

Summary of backward elimination								
Step	**Var. entered**	**Var. removed**	**Number of vars.**	**Partial R^2**	**Model R^2**	**C_p**	**F Value**	**Pr > F**
1		S	7	0.0054	0.496	7.669	0.670	0.417
2	V		8	0.0054	0.501	9.000	0.670	0.417
3		J	7	0.0016	0.499	7.200	0.200	0.656
4		D	6	0.0059	0.494	5.932	0.740	0.393
5		F	5	0.0079	0.486	4.916	1.000	0.321
6		H	4	0.0085	0.477	3.968	1.070	0.305
7		V	3	0.0072	0.470	2.857	0.900	0.345

4.4.2.3. Forward Selection Procedure

The procedure first selects the variable most highly correlated with the dependent variable (R), and then calculates the overall F test for the regression line with the first variable added. If the regression is significant, the F statistic associated with the remaining variables is determined. Then, the procedure focuses on the largest partial F statistic, and tests its significance to add the new variable to the model. The procedure was stopped once there were no more significant variables to be added (Kleinbaum et al.

2008). This procedure used a confidence level of α=0.05. Table 4.17 summarizes the procedure.

Table 4.17. Summary of forward selection.

Summary of forward selection							
Step	**Variable entered**	**Number vars. In**	**Partial R^2**	**Model R^2**	**C_p**	**F Value**	**Pr > F**
1	A	1	0.1844	0.1844	34.3461	15.60	0.0002
2	Y	2	0.1118	0.2962	22.4516	10.80	0.0016
3	G	3	0.1738	0.4700	2.8570	21.97	<.0001

4.4.2.4. Stepwise Regression Procedure

This procedure is a modified forward selection regression that allows the reexamination of the variables, at every step, of the variables incorporated in the model in previous steps. Variables entered in previous steps might become insignificant with the inclusion of newer variables because of its relationship. This behavior is checked with a partial F test for each variable currently in the model, as though it is the most recent variable included, irrespective of its actual entry point into the model. The variable with the smallest non-significant partial F statistic is removed; the model is refitted with the remaining variables, the partial F statistics are obtained and similarly examined; and so on. The procedure stops when no more variables can be added or removed (Kleinbaum et al. 2008). For this procedure a confidence level of α=0.05 was used for entry and removal criteria. Two stepwise procedures were conducted. The first procedure starts with no variables forced in the model. The second has three (A, Y, and G) fixed variables that are previously specified to be included in the model. Tables 4.18 and 4.19 show the summary of both procedures.

Table 4.18. Summary of stepwise selection.

Summary of stepwise selection								
Step	**Variable entered**	**Variable removed**	**Number vars. In**	**Partial R^2**	**Model R^2**	**C_p**	**F Value**	**Pr > F**
1	A		1	0.1844	0.1844	34.3461	15.60	0.0002
2	Y		2	0.1118	0.2962	22.4516	10.80	0.0016
3	G		3	0.1738	0.4700	2.8570	21.97	<.0001

Table 4.19. Summary of stepwise selection (Fixed variables A,Y, and G).

Summary of Stepwise Selection								
Step	**Variable entered**	**Variable removed**	**Number vars. In**	**Partial R^2**	**Model R^2**	**C_p**	**F Value**	**Pr > F**
1	AG		4	0.048	0.518	1.814	6.56	0.013

4.4.3. Summary of Model Selection Procedures

The results obtained from the different procedures vary according to inherit selection and elimination criteria of variables to be included in the model. The three methods used for the model selection have as result the same independent variables in the model (Y, A, and G). Therefore, the following step includes a stepwise procedure to include any potential interaction parameter that could be beneficial for the estimation of the dependent variable. It was determined that the interaction term AG did contribute significantly to the model.

4.4.4. Final Model Selection

In conclusion, the model selected includes the variables Y, A, and G. The model includes the variables that were the most significant, relevant, and representative of the characteristics of traffic and red light running behavior in the locations studied. Although, it was determined that the interaction parameter GA contributed significantly to the

model, serious collinearly was generated. Therefore, the model includes only the three basic variables without any interaction term. The model parameter estimates are listed in Table 4.20 and the model equation E4.1.

Table 4.20. Parameter estimates.

Variable	Parameter estimate
Intercept	79.15119
Y	0.47480
A	-0.08344
G	0.13212

$$\boldsymbol{R} = 79.15119 + 0.47480\,\boldsymbol{Y} - 0.08344\,\boldsymbol{A} + 0.13212\,\boldsymbol{G}$$
$$(R^2 = 0.470, Adj.R^2 = 0.446)$$

(E 4.1.)

The variable "A" is considered an important variable to be included in the final model because of its magnitude on the overall traffic volume, the interaction with other vehicles, and the yellow running. It was observed during the collection of the data that small vehicles tend to incur in red light running violations when larger vehicles were present since they provide a sense of coverage of any potential collisions which provides a sense of protection and security to violate the red light. Also, small vehicles have better speed development and reaction, so it is expected that their behavior would be prone to be more aggressive under the traffic light control.

The second independent variable in the model also represents a vehicle type. "G" plays an important role in the model since it represents single unit trucks that have significant presence in the overall traffic flow as well. In the literature, some studies included the influence of trucks to predict the red light running. Therefore, the variable is

a significant contribution to the prediction of red light running in this study. The contribution can be justified with the vehicle, operation, and driver characteristics. Since trucks have larger dimensions, drivers can find use it as an advantage and intimidate other vehicles in the intersection while violation the signal. In case of an incident, trucks would be the less affected. Additionally, the operational characteristics play an important role because this vehicles transport heavy cargos and the stopping capability is reduced. Therefore, they might be prone to violate the signal whenever they are surprised by the sudden signal change. Regarding driver behavior, there is a particular trend while discussing the behavior of truck drivers in Bolivia in general. Most of this drivers travel mostly in rural areas where little or no signals are present. They are reckless when they transition from the countryside roads to urban areas. Therefore, the drivers tend to disobey signals.

Finally, the variable "Y" is also included in the model. The yellow light running rate has some correlation with both variables already included in the model. Since both vehicles present the most significant red light running behavior, the yellow violation is significant because of its relation with the dependent variable. There is a similarity between the yellow and red light running since both are strictly related to the opportunity of occurrence of both variables.

4.5. Comparison of Regression Models

The selection of the model variables in the previous sections was conducted with all the data available from the three case studies. They were not separated or analyzed by intersection or approach. Therefore, the comparison of regression lines was used to

determine if the model's independent variables from each location significantly contributes to the overall model, or there were significant differences among the intersections that the overall model cannot be used as general predictor.

4.5.1. Dummy Variables

The dummy variables introduced in the model are displayed in Table 4.21, and equations E 4.2 to E 4.8 represent each intersection prediction model of the red light running with the already selected independent variables. The following equations represent the model for each case, and their parameters were tested in the next section to verify parallelism and coincidence among them.

$$W = \begin{cases} 1, & if\ \boldsymbol{Peru}\ intersection \\ 0, & otherwise \end{cases}$$

$$Z = \begin{cases} 1, & if\ \boldsymbol{Beijing}\ intersection \\ 0, & otherwise \end{cases}$$

Table 4.21. Dummy variables

Intersection	W	Z
Pablo	0	0
Beijing	0	1
Peru	1	0

- **General**

$$R = \beta_0 + \beta_1 Y + \beta_2 A + \beta_3 G + \beta_4 W + \beta_5 Z + \beta_6 YW + \beta_7 AW + \beta_8 GW + \beta_9 YZ + \beta_{10} AZ + \beta_{11} GZ + E \quad \text{(E 4.2.)}$$

- **Pablo**

$$R\begin{pmatrix} W = 0 \\ Z = 0 \end{pmatrix} = \beta_0 + \beta_1 Y + \beta_2 A + \beta_3 G + E \quad \text{(E 4.3.)}$$

$$R = 12.31177 - 0.19119\, Y + 0.04116\, A + 0.01237\, G \quad \text{(E 4.4.)}$$

$R^2 = 0.0523, Adj.R^2 = -0.053$

- **Beijing**

$$R\begin{pmatrix} W=0 \\ Z=1 \end{pmatrix} = (\beta_0 + \beta_5) + (\beta_1 + \beta_9)Y + (\beta_2 + \beta_{10})A + (\beta_3 + \beta_{11})G + E \quad \text{(E 4.5.)}$$

$$R = 33.96437 + 0.19197\,Y - 0.03050\,A + 0.04677\,G \quad \text{(E 4.6.)}$$

$R^2 = 0.2174, Adj.R^2 = 0.1$

- **Peru**

$$R\begin{pmatrix} W=1 \\ Z=0 \end{pmatrix} = (\beta_0 + \beta_4) + (\beta_1 + \beta_6)Y + (\beta_2 + \beta_7)A + (\beta_3 + \beta_8)G + E \quad \text{(E 4.7.)}$$

$$R = -25.36947 + 0.06830\,Y + 0.03102\,A + 0.10038\,G \quad \text{(E 4.8.)}$$

$R^2 = 0.1882, Adj.R^2 = -0.015$

4.5.2. Overall Hypothesis Testing: Parallelism and Coincidence

Test involving parallelism and coincidence among the regression models was performed. The null and alternate hypothesis for both test are:

- **Parallel**

  H_o: $\beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

  H_A: $Lines\ are\ not\ parallel$

- **Coincident**

H_o: $\beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

H_A: $Lines\ are\ not\ coincident$

The results from the test support that the regression models were parallel among themselves, but they were not coincident as shown in Table 4.22. Pairwise comparisons were conducted to determine which intersection model differs from the group.

Table 4.22. Parallel and coincident tests all regression models.

Test	P-value	Hypothesis
Parallelism	0.7898	Do not reject Ho. There is evidence that the models are parallel
Coincidence	< 0.0001	Reject Ho. There is evidence that the models are not coincident

4.5.3. Pairwise Hypothesis Testing: Parallelism and Coincidence

- **Pablo-Beijing**

H_o: $\beta_9 = \beta_{10} = \beta_{11} = 0$

H_A: $Lines\ are\ not\ parallel$

H_o: $\beta_5 = \beta_9 = \beta_{10} = \beta_{11} = 0$

H_A: $Lines\ are\ not\ coincident$

Table 4.23. Parallel and coincident tests Pablo-Beijing.

Test	P-value	Hypothesis
Parallelism	0.5081	Do not reject Ho. There is evidence that the models are parallel
Coincidence	< 0.0001	Reject Ho. There is evidence that the models are not coincident

- **Pablo-Peru**

H_o: $\beta_6 = \beta_7 = \beta_8 = 0$

H_A: $Lines\ are\ not\ parallel$

H_o: $\beta_4 = \beta_6 = \beta_7 = \beta_8 = 0$

H_A: $Lines\ are\ not\ coincident$

Table 4.24. Parallel and coincident tests Pablo-Peru.

Test	P-value	Hypothesis
Parallelism	0.8662	Do not reject Ho. There is evidence that the models are parallel
Coincidence	0.3125	Do not reject Ho. There is evidence that the models are coincident

- **Beijing-Peru**

$H_o: \beta_6 = \beta_9, \beta_7 = \beta_{10}, \beta_8 = \beta_{11}$ $H_o: \beta_4 = \beta_5, \beta_6 = \beta_9, \beta_7 = \beta_{10}, \beta_8 = \beta_{11}$

H_A: $Lines\ are\ not\ parallel$ H_A: $Lines\ are\ not\ coincident$

Table 4.25. Parallel and coincident tests Beijing-Peru.

Test	P-value	Hypothesis
Parallelism	0.3392	Do not reject Ho. There is evidence that the lines are parallel
Coincidence	0.1539	Do not reject Ho. There is evidence that the lines are coincident

The coincident test between Pablo and Beijing intersection was the only pairwise test that it was not significant with a P value of < 0.0001. The rest of the combinations were significant, so these regressions models were coincident and parallel at a significance level of 0.05.

4.5.4. Conclusions of Comparison of Regression Models

The regression models were parallel among them. Regarding coincidence, the only regression models that were not the same were Pablo and Beijing. However, a review of the results from the coincident test was conducted, which consisted basically on the analysis of the intercept values in the models. The range of difference between the intercepts of the models was significant, but there are not extreme differences that could

generate a completely biased estimate. It was concluded that despite the lack of significant coincidence between these two regression models representing two different intersections, the model had a fair fit over the overall sample. Therefore, the general model was assumed to be the most appropriate regression model for the prediction of red light running violation at the three locations under study.

4.6. Model Diagnostics

This sections focuses on verifying that the linear regression assumptions have been met, problems with collinearity, and potential outliners from the selected model including the variables R=Y A G.

4.6.1. Linear Regression Assumptions: Homoscedasticity and Independence

An overview of the Jackknife Residual plot in Figure 4.7 is performed, and the plot appears to reflect random scattering around the value of zero, so that no regression assumptions violations are apparent for the overall model.

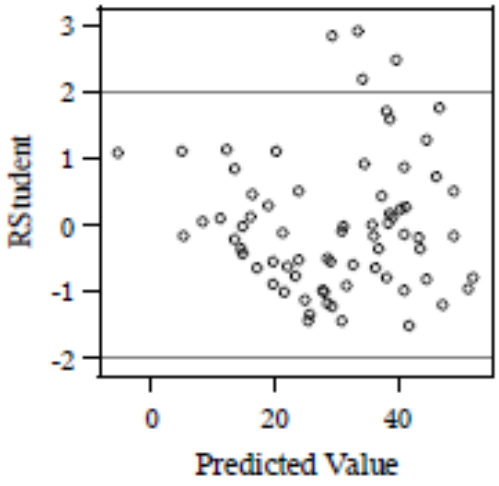


Figure 4.7. Jackknife residual plot.

4.6.2. Normality Assumptions

To assess the normality assumptions, the statistics such as the Skewness and Kurtosis were considered. The normal probability plot and QQ-plot were examined, and a measure of goodness of fit such as the Kormologorov-Smirnov test was performed.

The Skewness (0.9173) did not differ considerably from the ideal value of zero, and the Kurtosis (0.5469) did not have an extreme value and was not larger than 3 (heavier tail). Both values were not large statistics that displayed any significant normality assumption violation.

The normal probability plot in Figure 4.8 has a fairly normal distribution shape. Also, the QQ-plot in Figure 4.9 shows a fairly linear distribution of the normal percentiles. The goodness of fit test, the Kormologorov-Smirnov test, showed the distribution was not different from normal (P-value 0.0394).

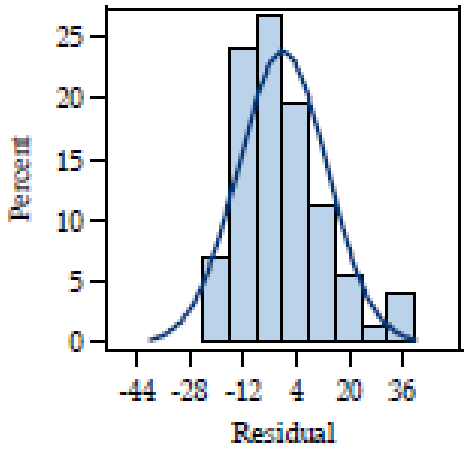


Figure 4.8. Residual probability distribution.

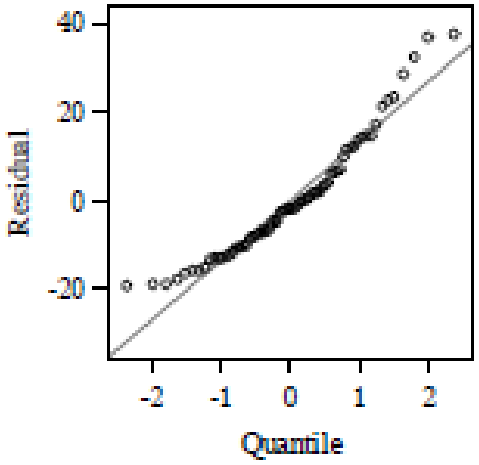


Figure 4.9. QQ-plot.

In conclusion, by determining and observing the different normal related statistics, plots, and test results, no apparent violation of the normal assumptions is observed.

4.6.3. Outliners

The potential outliners were identified by using Cook's Distance (di>1), Leverage statistics (hi>0.225), and Jackknife Residuals ($|r(-1)| > t_{64,0.975}=1.998$). Observations 50, 51, 65, and 69 were identified as potential outliners. It was determined that a particular pattern of these observations occurred at the noon or morning peak hour, and these are variables that have a high increase that generates a considerable influence in the R and Y variables. During the peak hours, congested conditions were found, and drivers at those times were returning to their homes lunch or end of the working day schedule. Higher signal light running were recorded during those periods because of aggressive and impatient driving behavior. Therefore, these outliners are plausible.

4.6.4. Collinearity

Collinearity exists when there are strong linear relationships among dependent variables (Kleinbaum et al. 2008). In order to observe these linear relationships, a correlation among the variables in the model was conducted in Table 4.26.

Table 4.26. Correlation of model variables.

Variable	Y	A	G	R
Y	1.0000	0.4873	-0.3113	0.0827
A	0.4873	1.0000	-0.2419	-0.4294
G	-0.3113	-0.2419	1.0000	0.4236
R	0.0827	-0.4294	0.4236	1.0000

Additionally, the variance inflation factors (VIF) were examined for each predictor in the model. The VIF values provide an indication of the degree to which the estimated standard errors of regression parameters estimates are affected by linear relationships among the predictor variables (Kleinbaum et al. 2008). As can be observed in Table 4.27, none of the independent variables show evidence of collinearity since their values are lower than 10. In the model selection procedure, the interaction parameter AG was significant and contributed to the prediction of the dependent variable. However, when the VIF was verified for such model structure (R=Y A G AG) severe collinearity problems were present. The VIF values were higher than a hundred for the interaction term and higher than ten for some of the basic variables. Therefore, it was discarded as the final model.

Table 4.27. Variance inflation factor of model variables.

Variable	**Variance inflation factor (VIF)**
Intercept	0
Y	1.383
A	1.327
G	1.120

4.7. Conclusion

The procedure to determine the best multivariable lineal regression model included different method of variable selection, comparison of regression models according to the data differences, and a thorough diagnostic of the model selected to verify the whole selection procedure. The variables considered to be the most relevant and significant for predicting red light running are: Y, A, and G. Additionally, while comparing regression models according to the specific locations and data, it was

determined that the models were parallel among themselves. It was also identified that there was significant difference between Beijing and Pablo, and they were not coincident. However, the models were assumed to be parallel and coincident despite of the particular difference because it was considered that the model was representative enough and adequate for the whole dataset in general. Regarding model diagnostics, the normality assumption was met. Some potential outliners were verified, and they were considered plausible because they are extreme values occurring at peak hours. Also, the correlations among dependent variables were not significant, and no collinearity problems were evident. Additional information regarding the independent variables selected for the model can be observed in Appendix B.

In conclusion, the general multivariable lineal model was considered the best fit for the prediction of red light running over the locations under study.

$$\boldsymbol{R} = 79.15119 + 0.47480\,\boldsymbol{Y} - 0.08344\,\boldsymbol{A} + 0.13212\,\boldsymbol{G}$$
$$(R^2 = 0.470, Adj.R^2 = 0.446)$$

(E 4.1.)

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1. Overview

Traffic circle deficiencies contribute to poor operational performance and safety. For this research, the study of red light running was considered. Initially, related literature review was examined to acknowledge the information available. Based on such background, a methodology for the collection of the data was developed. Video recordings of three traffic circles in Cochabamba, Bolivia were conducted. Additionally, a video processing method was also developed for the recordings collected using video screen display and keyboard keys to represent each vehicle type and movement simultaneously. Once the data was stored and tabulated, the analysis of the different available parameters involving red light running were conducted. The most important findings are summarized in the next section.

5.2. Summary of Findings

In this section, the most significant findings of the red light running analysis are summarized.

5.2.1. Red Light Running Related to Approaching Speed

The Peru intersection is used as a control location for the analysis of approaching speed on red light running. The approaches present particular differences that contribute to a significant speed differential, and the magnitude of speed was taken into account for the analysis. The hypothesis test supports that the occurrence of red light running at the particular intersection had significant higher rates at the west approach where higher

speeds were recorded. Therefore, higher red light running rates are found with higher approaching speeds.

5.2.2. Implementation of Countdown Panel to Reduce Red Light Running

Two approaches were used to analyze the contribution of counter down panels: short and long term performance. Initially, the analysis focuses on Pablo intersection to determine the benefit of implementing the device in the short term. A before and after analysis was conducted, and the time frame between the collection of the data was 6 weeks. Consequently, with the implementation of the countdown panels at Pablo intersection, there was significant reduction in red light running violations on a short-term basis. The reduction of red light running after the implementation registered a similar reduction of 31.5% violations at both approaches.

On a long term basis, Beijing intersection was used as a control location to assess the contribution of counter down panel implementation. In this case, a different approach was conducted. The east approach has had the counter down panel for over five months, and the west approach has never had the device. The reasons of such control traffic configuration are unknown. Finally, after the corresponding analysis, red light running rates were not significantly different at both approaches. Therefore, there was not a significant contribution of countdown panels to reduce red light running violations on a long term basis.

On a timely basis, the effectiveness of counter down panels was performed using two different intersections with distinct characteristics. There is the possibility that the results were influenced by each intersection configuration. However, there is a strong

believe that the results are relevant because of driver behavior adaptability and aggressiveness.

Regarding the influence of the counter down panels on the approaching speed, it was determined that there were significant differences between before and after implementation. But the magnitude of the differences were minimal, being between 0.8% to 1.3% with the short-term implementation.

5.2.3. Geometric Configuration: Traffic Circle to Conventional Intersection

The analysis of geometric configuration focuses on determining the changes on red light running rates when a former traffic circle was changed to a conventional four legged intersection. The approach of the analysis used Pablo intersection (former traffic circle to a four legged intersection) and Peru/Beijing (traffic circles) as control sites. Since the difference between the two group rates is evident, no statistical test was performed to determine its significance. Therefore, an average 28.70 (RLR/hr) (156%) of increase in red light running rates were found from a traffic circle to a conventional four leg intersection under similar roadway, traffic signal control, and vehicle configuration.

Regarding the approaching speed, a significant increase of 3.3% (2.1 Km/h) is found to be generated by the geometric change configuration.

5.2.4. Model Development

A multivariable linear regression model was determined for estimation red light running rates at traffic circles for the main approaches. The model form is displayed in equation E 4.1. The most significant variables included in the model were Y (yellow

running rate), A (traffic flow of small vehicles), and G (traffic flow of single unit trucks). Some analysis in order to include interaction terms in the model were conducted, and the most significant term was AG. The interaction term proved to contribute to the prediction of the dependent variable and the model in general; however, it was not considered or included in the final model because its inclusion generated serious collinearity.

$$\boldsymbol{R} = 79.15119 + 0.47480\,\boldsymbol{Y} - 0.08344\,\boldsymbol{A} + 0.13212\,\boldsymbol{G}$$
$$(R^2 = 0.470, Adj.R^2 = 0.446)$$

(E 4.1.)

5.3. Results

Red light running proved to be a significant and troublesome concern at the region of study. Not only does introduce safety concerns but also diminishes operational capacity considerably. Therefore, the current implemented countermeasures and conditions were studied to determine the most appropriate conditions to improve safety and operation of the system by reducing red light running rates.

The influence of traffic flow and approaching speed play an important role in red light running behavior. The findings of this research suggest that red light violations increase as both (flow or speed) increase. Similar research found similar evidence, and they related this relationship to the opportunities generated because of a rise in traffic flow, and the probability of stopping due to different approaching speeds.

The implementation of countdown panels in the traffic signal control did contribute to the reduction of red light running on a short time basis only. The results show significant reduction in red light signal violation when the device was just

implemented, but as time passed by; the rates went back to normal. The countdown panel lost its effectiveness in time. It was assumed that such trend occurs due to driver adjustment to the signal device implementation. Regular commuting drivers would evidently feel the change more significant at the beginning of the implementation, but they would get enough experience and find the appropriate time to beat the signal. In this case, it would result in drivers increasing the speed at a determined distance range to the intersection once they see a specific time remaining in the signal.

A geometric configuration from a traffic circle to a four leg conventional intersection was found to be a poor countermeasure to safety and operational issues of traffic circles. There was significant increase on red light running rates and approaching speeds. Additionally, the geometrics at the particular site location did not provide enough guidance for drivers and pedestrians to use the facility in a safe and efficient manner. Therefore, the countermeasure was considered to be a setback, and it is not recommended for future projects in the corridor nor the region.

Regarding model development, it was generated to provide guidance and estimation of red light running rates for future assessments of incidents and operation purposes. The variables in the model are easily obtainable from field data collection, and they can contribute to planning and organizational purposes. From the literature review, some authors also considered that the presence of heavy vehicles in their models play a significant role in the occurrence on red light running, and in this research the same criteria was found. More specifically, as it was explained in the general considerations and site locations description, the corridor presents an important presence of heavy vehicles using not only the corridor but also the particular intersections, and its inclusion

in the model supports the observed behavior in the field along with the other independent variables.

5.4. Recommendations

The following suggestions are stated for the region of study:

- The current signal configurations of the traffic circles provide signal devices at the entrance and exit of the approach which generates more confidence on drivers to negotiate through the intersection violating the signals. Drivers tend to keep moving from the first signal to the second whenever they have the opportunity, and pedestrians are severely endangered because drivers are not only focusing on the conflictive traffic but also on the signal. This situation occurs when pedestrians have the right of way and vehicles are supposed to completely stop while the signals displays red at the entrance to the intersection. Therefore, it is suggested to avoid providing multiple traffic signal displays in order to commit drivers to comply with the main signal at the entry of the intersection.
- The implementation of countdown panels proved to be an efficient countermeasure on a short term basis only. However, during the data collection, several other benefits have been observed form the device. The counter provides significant guidance to traffic and pedestrians during highly congested scenarios. It helps to hold vehicles from invading the intersection when the circulatory roadway is already saturated and there is no room left to enter. As a consequence, drivers are provided important information to decide under such conditions whether to enter or hold back at an appropriate location in order not to get stuck in the middle of the conflicting traffic

roadway. Therefore, its implementation is encouraged for the intersections in the corridor under study.

- The addition of more time in the yellow phase signal timing is suggested. It would contribute to the reduction of red light running behavior. The current yellow signal time interval is 3 seconds in all intersections.
- An all red interval is also suggested to be implemented as part of the traffic signal control timing. In this case, traffic circles accumulate vehicles in the circulatory roadway, and during congested conditions the queues invade the conflicting traffic roadway. Therefore, an addition of all red signal timing would allow vehicles to clear the paths of the entering vehicles to avoid locking up the approaches and the intersection in general.
- The previous suggestions have been based on low cost countermeasures or implementations. However, the benefit of such contributions is limited since their efficiency would be reduced as demand increases and drivers get adjusted to the changes with time. Therefore, from previous research studying a similar circumstance (Claros and Gonzales 2008) conducted and analysis of a major traffic circle geometric configuration. It consists of keeping the traffic circle at ground level separate the main traffic to cross the intersection by an under pass. The study found that the operational and safety benefits were significantly increased. Therefore, as a long term countermeasure project, the separation of the main flow from the conflicting traffic is suggested as a solution to overall safety and operation problems at conventional traffic circles.

REFERENCES

Baguley, C. J. (1988). "Running the red at signals on high-speed roads." Traffic Engineering & Control. Crowthorne, England, pp.415-420.

Bonneson, J., and Son H. (2003). "Prediction of expected red light running frequency at urban intersections." In *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., No.1830, pp. 38–47.

Bonneson, J., and Zimmerman K. (2001). "Review and evaluation of factors that affect the frequency of red-light-running." Federal Highway Administration. Publication FHWA/TX-02/0-4027-1. Washington, D.C.

Claros, B. R., and Gonzales A. G. (2008). "Proyecto de grado: Diseño distribuidor vehicular rotonda administradora boliviana de carreteras (A.B.C.)." Universidad Mayor de San Simon, Cochabamba, Bolivia.

Elmitiny, N., Yan, X., Radwan E., Russo, C., and Nashar D. (2010). "Classification analysis of driver's stop/go decision and red-light running violation." Accident Analysis and Prevention Elsevier, No. 42, pp. 101-111.

Farraher, B. A., Weinholzer R., and Kowski M. P. (1999). "The effect of advanced warning flashers on the red light running – A study using motion imaging recordings system technology at Trunk Highway 169 and Pioneer Trail in Bloomington, Minnesota" *Compendium of Technical Papers for the 69th Annual ITE Meeting*. (CD-ROM) Institute of Transportation Engineers, Washington, D.C.

Grant, G. S., and Eric S. T. (2009). "Advance warning signals – Long-term monitoring results." *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., No. 2122, pp. 27–35.

Grant, G. S., Peterson R., Eggett D. L., and Bradley C. G. (2007). "Effectiveness of blank-out overhead dynamic advance warning signals at high-speed signalized intersections." *Journal of Transportation Engineering*. ASCE, DOI: 10.1061/(ASCE)0733-947X(2007)133:10(564), pp. 564-571.

ITE Technical Council Task Force 4TF-1 (1994). "Determining vehicle signal change and clearance intervals." *An Informational Report of ITE*, Washington, D.C.

Kleinbaum, D., Kupper L., and Nizam A. (2008). *Applied Regression Analysis and Other Multivariate Methods*, 4th Ed., Thomson Brooks/Cole, Belmont, CA.

Mohamedshah, Y. M. Chen L.W., and Council F. M. (2000). "Association of selected intersection factors with red light running crashes." *Proceedings of the 70th Annual ITE Conference*. Institute of Transportation Engineers, Washington, D.C.

Persaud B., Hauer E., Retting R., Vallurupalli R., and Mucsi K. (1997). "Crash reductions related to traffic signal removal in Philadelphia." Accident Analysis and Prevention. Insurance Institute for Highway Safety, Arlington, Va., Vol. 29, No. 6, pp. 803-810.

Porter, B. E., and England K. J. (2000). "Predicting Red-Light Running Behavior: A Traffic Safety Study in Three Urban Settings." *Journal of Safety Research*. National Safety Council and Elsevier Science, Vol. 31, No. 1, pp. 1–8.

Retting, R. A., Williams A. F., and Greene M. A. (1998). "Red-light running and sensible countermeasures, summary of research findings." *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., No. 1640, pp. 23–26.

Schattler, K. L., Datta T. K., and Hill C. L. (2003). "Change and clearance interval design on red-light running and late exits." *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., No. 1856, pp. 193–201.

Tullio, G., and Savino R. (2006). "Evaluation of proneness to red light violation." *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, Washington, D.C., No. 1969, pp. 35–44.

Van der Horst, R., and Wilmink A. (1986). "Drivers' decision-making at signalized intersections: an optimization of the yellow timing." Traffic Engineering & Control, Crowthorne, England, pp. 615–622.

Yiyi, C., Xuesong W., and Xiaohong C. (2009). "Investigation of the relationship between red-light-running frequencies and intersection features." ICCTP 2009: Critical Issues in Transportation Systems Planning, Development, and Management, pp. 769-776.

APPENDIX A:HYPOTHESIS TESTING

A.1. Comparisons Dataset

Int.	Pablo								Beijing				Peru			
App.	ENC		WNC		EC		WC		EC		WNC		E		W	
N	Y	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	R	Y	R
1	24	40	56	68	52	16	64	60	20	16	51	12	95	33	132	37
2	16	64	52	48	44	28	40	68	59	0	67	16	66	33	66	33
3	40	40	52	52	44	64	64	48	55	0	24	8	58	29	83	37
4	48	44	64	76	88	48	36	64	39	24	39	12	83	29	116	33
5	36	48	68	72	56	36	48	28	32	16	43	16	62	12	70	25
6	36	84	60	56	52	20	48	32	36	12	55	8	66	37	87	37
7	40	60	72	80	32	36	48	36	39	16	51	8	41	8	99	17
8	32	60	32	76	56	76	52	32	32	0	43	20	83	12	54	17
9	48	80	60	68	28	68	88	40	39	8	32	12	116	25	66	33
10	36	44	64	60	68	48	56	28	36	12	47	16	50	12	70	17
11	48	52	84	68	80	44	40	60	39	16	59	20	50	29	70	29
12	72	88	72	52	60	28	60	36	36	0	51	4	62	17	62	17
13	48	28	64	60	44	40	68	52	43	8	83	8	66	25	116	33
14	60	32	56	76	72	24	72	32	39	8	32	4	87	17	87	17
15	68	48	24	52	48	20	64	64	39	8	51	12	87	21	91	33
16	52	68	74	92	80	28	56	28	47	11	47	0	79	21	99	21
17	52	36	96	68	16	28	68	40	55	20	67	20	107	31	123	38
18	40	36	52	40	20	32	52	40	55	0	39	8	58	12	83	45
19	64	72	96	60	44	24	52	28	43	20	47	0	79	8	103	21
20	64	80	80	56	84	28	40	56	36	12	43	28	54	12	91	25
21	32	40	88	68	48	52	76	60	32	4	8	32	91	41	79	29
22	44	64	48	48	36	24	72	24	28	12	55	95	95	17	79	21
23	36	20	84	104	24	32	56	44	51	8	39	28	70	8	99	29
24	44	32	84	68	20	28	80	28	47	12	28	20	58	17	95	29
25	36	68	68	32	36	28	76	52	28	4	24	8	103	17	79	17
26	44	32	76	32	44	24	88	40	20	8	16	8	95	17	91	29
27	60	40	68	52	64	44	72	32	43	12	51	4	99	37	95	29
28	32	72	68	48	40	40	96	32	39	8	36	12	112	29	91	17
29			100	44	40	32	56	20	43	4	36	16	141	8	74	29
30			80	28	16	16	84	32	47	0	36	4	103	17	103	45
31			48	21	64	40	52	12	32	8	59	8	124	25	103	21
32					64	56	72	44	58	0	58	5	74	12	95	12
33									87	16	79	55	99	8	66	25
34									32	20	55	24	161	15	107	46
35									47	12	59	4				
36									47	12	63	4				
37									55	12	39	8				
38									36	0	39	16				
39									47	8	91	4				
40									55	20	47	8				
41									87	20	63	16				
42									36	16	43	16				
43									28	8	32	8				
44									36	8	55	12				
45									55	4	67	24				
46									67	16	51	20				
47									47	16	83	32				
48									53	53	68	0				

A.2. Model Development Dataset

N	Inter.	W	Z	Count	V	J	A	S	D	F	G	H	Y	R	N	Inter.	W	Z	Count	V	J	A	S	D	F	G	H	Y	R
1	PERU	1	0	Peru E1	1576	74	1205	121	111	0	65	0	78	31	37	BEIJING	0	1	Beijing W3	2012	76	1235	421	89	8	166	18	64	22
2	PERU	1	0	Peru E1	1510	93	1121	129	105	0	61	1	64	17	38	BEIJING	0	1	Beijing W3	2160	83	1363	396	96	9	193	21	55	9
3	PERU	1	0	Peru E1	1601	122	1172	126	116	0	64	0	67	20	39	BEIJING	0	1	Beijing W3	2087	124	1317	329	79	3	209	26	48	13
4	PERU	1	0	Peru E1	1810	218	1304	126	108	0	53	0	83	23	40	BEIJING	0	1	Beijing W3	1981	113	1240	330	78	6	187	27	66	20
5	PERU	1	0	Peru E2	1600	73	1220	141	123	0	43	0	70	18	41	PABLO	0	0	Pablo EC1	1805	51	1028	452	87	18	137	32	57	39
6	PERU	1	0	Peru E2	1526	99	1126	168	103	0	30	0	82	15	42	PABLO	0	0	Pablo EC1	1694	65	892	441	79	14	184	19	49	42
7	PERU	1	0	Peru E2	1621	134	1168	166	108	0	44	0	114	23	43	PABLO	0	0	Pablo EC1	1705	86	943	406	81	10	156	23	59	47
8	PERU	1	0	Peru E2	1791	203	1294	138	105	0	51	0	105	15	44	PABLO	0	0	Pablo EC1	1809	107	1065	357	81	8	166	25	61	28
9	PERU	1	0	Peru W1	1566	64	1306	63	69	35	26	4	99	35	45	PABLO	0	0	Pablo EC2	1552	69	900	352	87	4	112	28	41	28
10	PERU	1	0	Peru W1	1551	104	1233	70	53	52	40	0	79	23	46	PABLO	0	0	Pablo EC2	1469	82	894	276	77	1	124	15	32	34
11	PERU	1	0	Peru W1	1593	115	1257	71	56	50	39	6	72	25	47	PABLO	0	0	Pablo EC2	1469	82	886	276	68	2	134	21	46	34
12	PERU	1	0	Peru W1	1791	179	1400	69	55	52	31	6	97	27	48	PABLO	0	0	Pablo EC2	1587	105	976	283	78	3	124	18	46	36
13	PERU	1	0	Peru W2	1793	98	1188	336	60	33	71	8	88	30	49	PABLO	0	0	Pablo ENC1	1638	82	938	364	78	10	141	25	32	47
14	PERU	1	0	Peru W2	1819	98	1215	327	66	30	75	9	88	22	50	PABLO	0	0	Pablo ENC1	1626	71	956	350	83	13	136	17	36	63
15	PERU	1	0	Peru W2	1813	100	1205	334	64	29	73	8	91	31	51	PABLO	0	0	Pablo ENC1	1736	104	1056	354	85	6	106	25	51	66
16	PERU	1	0	Peru W2	1788	86	1178	345	59	44	65	12	92	25	52	PABLO	0	0	Pablo ENC2	1793	72	1035	392	96	18	157	23	57	44
17	BEIJING	0	1	Beijing E1	1685	99	1098	345	64	12	49	17	43	10	53	PABLO	0	0	Pablo ENC2	1599	59	914	356	82	6	152	30	55	56
18	BEIJING	0	1	Beijing E1	1643	116	1091	279	43	36	68	9	35	11	54	PABLO	0	0	Pablo ENC2	1618	83	905	347	82	4	164	33	39	39
19	BEIJING	0	1	Beijing E1	1786	137	1163	318	34	44	63	27	38	9	55	PABLO	0	0	Pablo ENC2	1724	121	960	364	86	4	164	25	43	53
20	BEIJING	0	1	Beijing E1	2113	234	1326	392	30	59	48	24	41	8	56	PABLO	0	0	Pablo WC1	2067	100	1147	477	89	6	236	12	51	60
21	BEIJING	0	1	Beijing E2	1904	91	1183	442	39	54	63	31	47	13	57	PABLO	0	0	Pablo WC1	1928	83	1115	394	97	3	208	28	49	32
22	BEIJING	0	1	Beijing E2	1722	106	1088	356	38	40	57	37	39	9	58	PABLO	0	0	Pablo WC1	1925	82	1129	402	84	4	192	32	61	41
23	BEIJING	0	1	Beijing E2	1646	102	999	365	38	36	63	43	33	8	59	PABLO	0	0	Pablo WC1	1900	103	1127	376	80	5	183	26	67	45
24	BEIJING	0	1	Beijing E2	2038	177	1224	451	58	40	59	29	43	3	60	PABLO	0	0	Pablo WC2	1956	87	1093	432	89	4	229	22	53	41
25	BEIJING	0	1	Beijing E3	1917	75	1254	370	56	36	114	12	53	15	61	PABLO	0	0	Pablo WC2	1965	82	1159	426	88	4	188	18	71	39
26	BEIJING	0	1	Beijing E3	1734	85	1109	348	59	7	117	8	48	10	62	PABLO	0	0	Pablo WC2	1953	85	1135	405	76	8	210	34	83	39
27	BEIJING	0	1	Beijing E3	1775	89	1124	347	76	4	120	15	46	13	63	PABLO	0	0	Pablo WC2	1855	108	1126	371	79	2	146	22	67	28
28	BEIJING	0	1	Beijing E3	2095	167	1353	384	80	3	97	10	55	20	64	PABLO	0	0	Pablo WNC1	2019	90	1160	429	88	4	222	26	56	61
29	BEIJING	0	1	Beijing W1	2064	102	1173	455	96	13	213	13	45	12	65	PABLO	0	0	Pablo WNC1	1943	102	1157	403	82	4	177	18	58	71
30	BEIJING	0	1	Beijing W1	1955	122	1183	338	88	10	191	22	48	13	66	PABLO	0	0	Pablo WNC1	1905	136	1090	385	90	8	177	19	70	62
31	BEIJING	0	1	Beijing W1	1836	114	1158	291	69	2	184	18	47	13	67	PABLO	0	0	Pablo WNC1	1726	99	950	394	85	2	168	28	52	70
32	BEIJING	0	1	Beijing W1	1780	103	1151	276	88	8	133	22	53	6	68	PABLO	0	0	Pablo WNC2	2027	98	1176	439	91	3	202	18	81	56
33	BEIJING	0	1	Beijing W2	2141	122	1290	405	75	4	205	40	49	14	69	PABLO	0	0	Pablo WNC2	1993	113	1181	416	85	5	174	19	76	72
34	BEIJING	0	1	Beijing W2	1673	60	977	321	96	5	184	31	33	43	70	PABLO	0	0	Pablo WNC2	1916	89	1175	371	68	5	184	24	70	41
35	BEIJING	0	1	Beijing W2	1839	95	1104	329	89	7	181	36	32	8	71	PABLO	0	0	Pablo WNC2	1887	132	1112	360	83	3	182	17	78	32
36	BEIJING	0	1	Beijing W2	1888	103	1204	304	84	4	162	27	45	8															

A.3. SAS Code

```
OPTIONS LS=78 PS=65 NODATE NONUMBER;
ODS PDF FILE='Speed z-test';
ODS TRACE ON;
ODS GRAPHICS ON;
TITLE 'Speed H. Testing';
PROC IMPORT OUT= WORK.ztest
            DATAFILE= "E:\Research\Data\speed_count.xlsx"
            DBMS=EXCELCS REPLACE;
     RANGE="'Speed Differential$'";
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
data one;
set ztest;
ods graphics on;
proc ttest data=one side=2;
class PABLO;
var PABLOS;
run;
ods graphics off;
ods graphics on;
proc ttest data=one side=2;
class PABLOE1;
var PABLOE1S;
run;
ods graphics off;
ods graphics on;
proc ttest data=one side= L;
class PABLOW1;
var PABLOW1S;
run;
ods graphics off;
ods graphics on;
proc ttest data=one side=2;
class BEIJING;
var BEIJINGS;
run;
ods graphics off;
ods graphics on;
proc ttest data=one side=L;
class PERU;
var PERUS;
run;
ods trace off;
ods graphics off;
ods pdf close;
```

A.4. SAS Output

Speed H. Testing

The TTEST Procedure

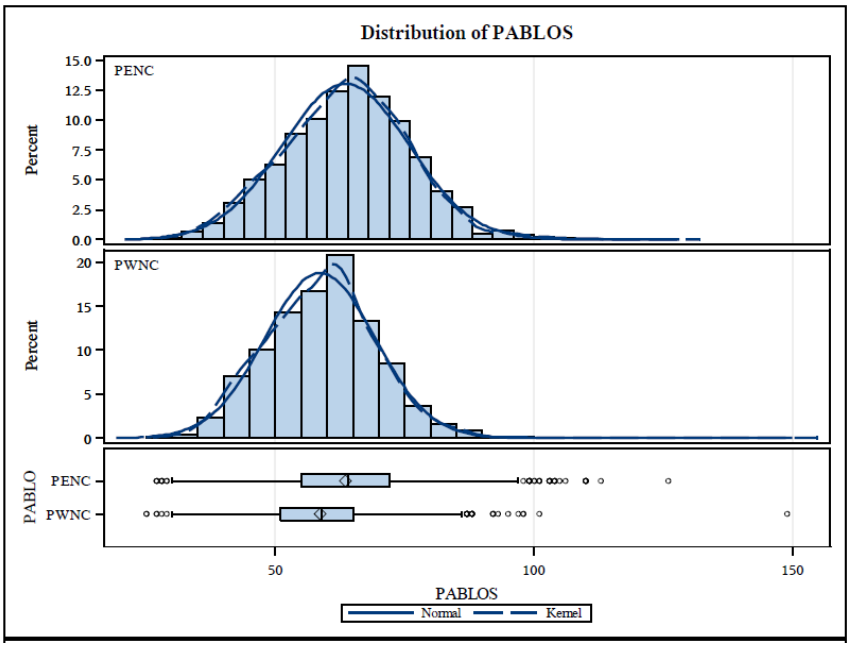
Variable: PABLOS (PABLOS)

PABLO	N	Mean	Std Dev	Std Err	Minimum	Maximum
PENC	5282	63.5757	12.2609	0.1687	27.0000	126.0
PWNC	4451	58.6594	10.6183	0.1592	25.0000	149.0
Diff (1-2)		4.9163	11.5388	0.2348		

PABLO	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
PENC		63.5757	63.2450	63.9065	12.2609	12.0315	12.4993
PWNC		58.6594	58.3474	58.9714	10.6183	10.4022	10.8436
Diff (1-2)	Pooled	4.9163	4.4561	5.3765	11.5388	11.3789	11.7032
Diff (1-2)	Satterthwaite	4.9163	4.4617	5.3710			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	9731	20.94	<.0001
Satterthwaite	Unequal	9723.7	21.20	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	5281	4450	1.33	<.0001



Speed H. Testing

The TTEST Procedure

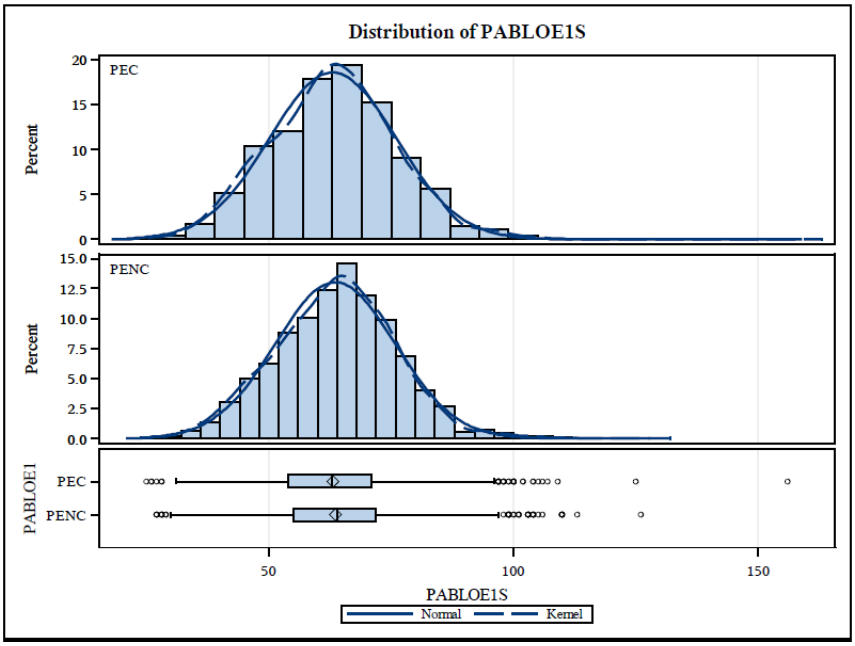
Variable: PABLOE1S (PABLOE1S)

PABLOE1	N	Mean	Std Dev	Std Err	Minimum	Maximum
PEC	2929	63.0662	12.8996	0.2384	25.0000	156.0
PENC	5282	63.5757	12.2609	0.1687	27.0000	126.0
Diff (1-2)		-0.5095	12.4925	0.2878		

PABLOE1	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
PEC		63.0662	62.5989	63.5336	12.8996	12.5776	13.2388
PENC		63.5757	63.2450	63.9065	12.2609	12.0315	12.4993
Diff (1-2)	Pooled	-0.5095	-1.0737	0.0547	12.4925	12.3043	12.6865
Diff (1-2)	Satterthwaite	-0.5095	-1.0820	0.0630			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8209	-1.77	0.0767
Satterthwaite	Unequal	5790.7	-1.74	0.0811

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2928	5281	1.11	0.0017



Speed H. Testing

The TTEST Procedure

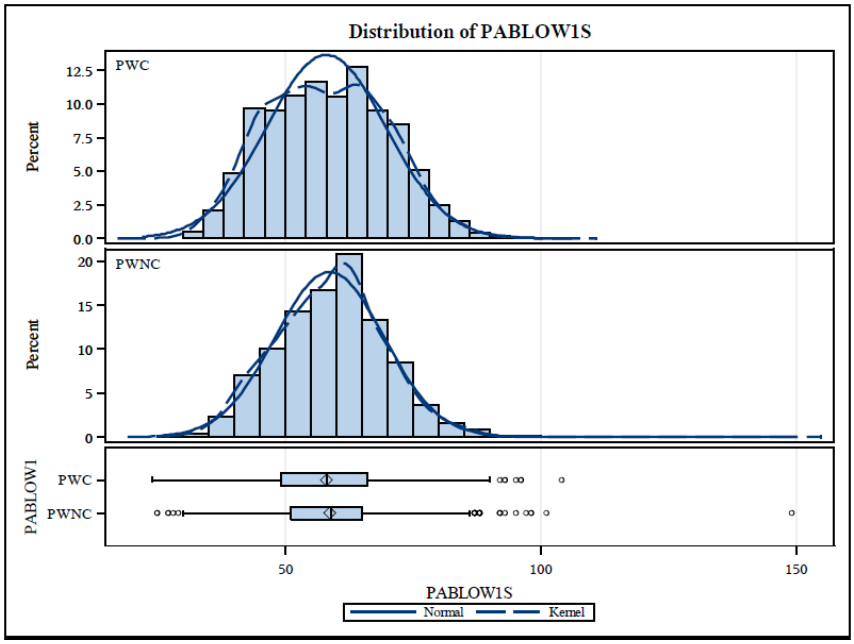
Variable: PABLOW1S (PABLOW1S)

PABLOW1	N	Mean	Std Dev	Std Err	Minimum	Maximum
PWC	3464	57.9925	11.6968	0.1987	24.0000	104.0
PWNC	4451	58.6594	10.6183	0.1592	25.0000	149.0
Diff (1-2)		-0.6669	11.1032	0.2516		

PABLOW1	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
PWC		57.9925	57.6028	58.3821	11.6968	11.4277	11.9789
PWNC		58.6594	58.3474	58.9714	10.6183	10.4022	10.8436
Diff (1-2)	Pooled	-0.6669	-Infty	-0.2531	11.1032	10.9329	11.2789
Diff (1-2)	Satterthwaite	-0.6669	-Infty	-0.2481			

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	7913	-2.65	0.0040
Satterthwaite	Unequal	7067.2	-2.62	0.0044

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3463	4450	1.21	<.0001



Speed H. Testing

The TTEST Procedure

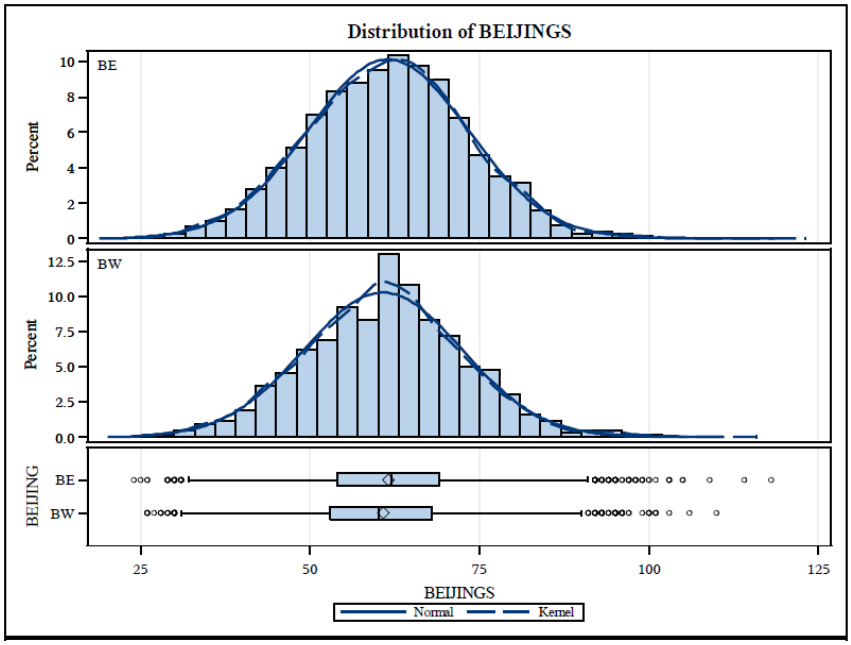
Variable: BEIJINGS (BEIJINGS)

BEIJING	N	Mean	Std Dev	Std Err	Minimum	Maximum
BE	6559	61.4934	11.8318	0.1461	24.0000	118.0
BW	5006	60.7549	11.6394	0.1645	26.0000	110.0
Diff (1-2)		0.7385	11.7489	0.2205		

BEIJING	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
BE		61.4934	61.2070	61.7798	11.8318	11.6328	12.0379
BW		60.7549	60.4324	61.0774	11.6394	11.4158	11.8720
Diff (1-2)	Pooled	0.7385	0.3063	1.1707	11.7489	11.5994	11.9023
Diff (1-2)	Satterthwaite	0.7385	0.3072	1.1697			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	11563	3.35	0.0008
Satterthwaite	Unequal	10858	3.36	0.0008

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6558	5005	1.03	0.2174



Speed H. Testing

The TTEST Procedure

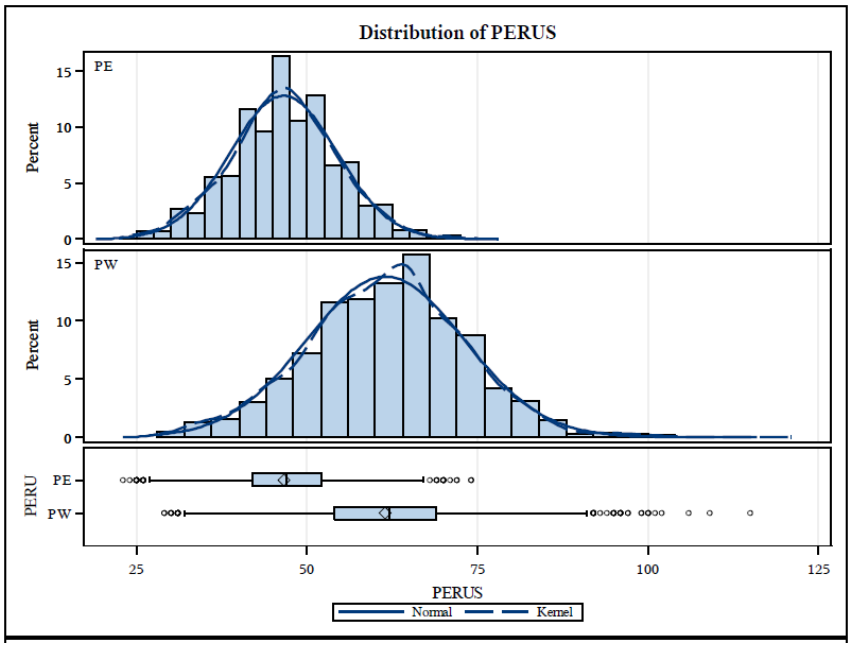
Variable: PERUS (PERUS)

PERU	N	Mean	Std Dev	Std Err	Minimum	Maximum
PE	2989	46.5544	7.7955	0.1426	23.0000	74.0000
PW	3517	61.4822	11.5544	0.1948	29.0000	115.0
Diff (1-2)		-14.9279	10.0044	0.2489		

PERU	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
PE		46.5544	46.2748	46.8339	7.7955	7.6027	7.9983
PW		61.4822	61.1002	61.8642	11.5544	11.2905	11.8309
Diff (1-2)	Pooled	-14.9279	-Infty	-14.5184	10.0044	9.8354	10.1794
Diff (1-2)	Satterthwaite	-14.9279	-Infty	-14.5307			

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	6504	-59.98	<.0001
Satterthwaite	Unequal	6198.6	-61.83	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3516	2988	2.20	<.0001



APPENDIX B: MULTIVARIABLE LINEAR REGRESSION MODEL

B.1. SAS code

```
ODS PDF FILE='RLR REGRESSION MODEL';
ODS TRACE ON;
ODS GRAPHICS ON;
OPTIONS LS=78 PS=65 NODATE NONUMBER;
PROC IMPORT OUT= WORK.regmodel1
DATAFILE= "E:\Research\Data\flow_red_yellow_count.xlsx"
DBMS=EXCELCS REPLACE;
RANGE="'regmodel(1)$'";
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;
PROC IMPORT OUT= WORK.REGMODEL2
DATAFILE= "E:\Research\Data\flow_red_yellow_count.xlsx"
DBMS=EXCELCS REPLACE;
RANGE="'regmodel(2)$'";
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;
PROC IMPORT OUT= WORK.REGMODEL3
DATAFILE= "E:\Research\Data\flow_red_yellow_count.xlsx"
DBMS=EXCELCS REPLACE;
RANGE="'regmodel(3)$'";
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;
PROC IMPORT OUT= WORK.REGMODEL4
DATAFILE= "E:\Research\Data\flow_red_yellow_count.xlsx"
DBMS=EXCELCS REPLACE;
RANGE="'regmodel(4)$'";
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;
RUN;
data one;
set regmodel1;
AS=A*S;
AD=A*D;
AG=A*G;
AH=A*H;
YA=Y*A;
YS=Y*S;
YD=Y*D;
YG=Y*G;
YH=Y*H;
YGA=Y*G*A;
title 'model selection';
Proc print data=one;
PROC REG DATA=one;
```

```
MODEL R=Y J A S D F G H V / SELECTION=ADJRSQ RSQUARE CP AIC BEST=200;
PROC REG DATA=one;
MODEL R=Y J A S D F G H V / SELECTION=STEPWISE SLE=0.05 SLS=0.05
DETAILS=ALL;
PROC REG DATA=one;
MODEL R= Y A G J S D F H AS AD AG AH YA YS YD YG YH YGA /
SELECTION=STEPWISE SLE=0.05 SLS=0.05 INCLUDE=3;
PROC REG DATA=one;
MODEL R=Y J A S D F G H V / SELECTION=FORWARD SLE=0.05 DETAILS=SUMMARY;
PROC REG DATA=one;
MODEL R=Y J A S D F G H V / SELECTION=BACKWARD SLS=0.05
DETAILS=SUMMARY;
DATA two;
SET one;
YW=Y*W;
AW=A*W;
GW=G*W;
AZ=A*Z;
YZ=Y*Z;
GZ=G*Z;
PROC SORT DATA=two;
BY W Z;
PROC REG DATA=two;
MODEL R=Y A G W Z YW AW GW YZ AZ GZ;
BY W Z;
PROC REG DATA=two;
MODEL R=Y A G W Z YW AW GW YZ AZ GZ;
PARALLEL: TEST YW=AW=GW=YZ=AZ=GZ=0;
COINCIDENT: TEST W=Z=YW=AW=GW=YZ=AZ=GZ=0;
title 'all';
RUN;
DATA three;
SET regmodel2;
YW=Y*W;
AW=A*W;
GW=G*W;
PROC REG DATA=three;
MODEL R=Y A G W Z YW AW GW;
PARALLEL: TEST YW=AW=GW=0;
COINCIDENT: TEST W=YW=AW=GW=0;
title 'peru-pablo';
RUN;
DATA four;
SET regmodel3;
AZ=A*Z;
YZ=Y*Z;
GZ=G*Z;
PROC REG DATA=four;
MODEL R=Y A G Z YZ AZ GZ;
PARALLEL: TEST YZ=AZ=GZ=0;
COINCIDENT: TEST Z=YZ=AZ=GZ=0;
title 'beijing-pablo';
RUN;
DATA five;
SET regmodel4;
YW=Y*W;
```

```
AW=A*W;
GW=G*W;
PROC REG DATA=five;
MODEL R= Y A G W YW AW GW;
PARALLEL: TEST YW=AW=GW=0;
COINCIDENT: TEST W=YW=AW=GW=0;
title 'peru-beijing';
RUN;
data six;
title 'Diagnostics';
set regmodel1;
PROC UNIVARIATE DATA=six NEXTROBS=5;
VAR R Y A G;
PROC REG DATA=six CORR;
MODEL R=Y A G/ INFLUENCE VIF;
OUTPUT OUT=seven P=yhat R=residual Rstudent=delStudentResidual;
PROC UNIVARIATE DATA=seven NORMAL NEXTROBS=5;
VAR residual delStudentResidual;
ID R Y A G;
RUN;
ODS TRACE OFF;
ODS GRAPHICS OFF;
ODS PDF CLOSE;
```

B.2. SAS output

B.2.1. Model Selection Procedures

- **All Possible Regressions Procedure**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R

Adjusted R-Square Selection Method

Number in Model	Adjusted R-Square	R-Square	C(p)	AIC	Variables in Model
3	0.4463	0.4700	2.8570	375.7256	Y A G
5	0.4461	0.4856	4.9149	377.6004	Y A D F G
6	0.4461	0.4936	5.9319	378.4999	Y A F G H V
5	0.4461	0.4856	4.9163	377.6019	Y A G H V
4	0.4459	0.4775	3.9219	376.7103	Y A F G
7	0.4456	0.5010	7.0071	379.4488	Y J A S F H V
4	0.4455	0.4772	3.9677	376.7603	Y A G V
5	0.4449	0.4845	5.0556	377.7565	Y A F G H
6	0.4440	0.4916	6.1712	378.7693	Y J A S H V
7	0.4438	0.4994	7.2002	379.6696	Y A D F G H V
6	0.4434	0.4911	6.2391	378.8456	Y J A D F G
6	0.4430	0.4907	6.2810	378.8927	Y A D F G V
7	0.4429	0.4986	7.3003	379.7838	Y J S D F G V
4	0.4427	0.4745	4.2942	377.1162	Y A S G
6	0.4417	0.4896	6.4294	379.0589	Y A D F G H
6	0.4410	0.4889	6.5092	379.1482	Y J A F G H
7	0.4409	0.4968	7.5236	380.0378	Y A S F G H V
6	0.4409	0.4888	6.5179	379.1579	Y A S D F G
5	0.4409	0.4808	5.5139	378.2627	Y A F G V
5	0.4408	0.4808	5.5231	378.2727	Y J A F G
6	0.4407	0.4886	6.5446	379.1878	Y A S F G H
5	0.4406	0.4805	5.5493	378.3015	Y A S G H
4	0.4404	0.4724	4.5619	377.4067	Y J A G
4	0.4398	0.4719	4.6288	377.4791	Y A G H
7	0.4397	0.4957	7.6673	380.2007	Y J A S D H V
7	0.4396	0.4957	7.6688	380.2025	Y J A D F G H
7	0.4392	0.4953	7.7190	380.2593	Y A S D F G H
6	0.4389	0.4870	6.7427	379.4088	Y A S G H V
5	0.4387	0.4788	5.7629	378.5361	Y A S F G
5	0.4384	0.4785	5.8051	378.5824	Y J A S V

- **Backward Elimination procedure**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Number of Observations Read	71
Number of Observations Used	71

Summary of Backward Elimination									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1		S	S	7	0.0054	0.4957	7.6688	0.67	0.4166
2	V		V	8	0.0054	0.5011	9.0000	0.67	0.4166
3		J	J	7	0.0016	0.4994	7.2002	0.20	0.6561
4		D	D	6	0.0059	0.4936	5.9319	0.74	0.3926
5		F	F	5	0.0079	0.4856	4.9163	1.00	0.3208
6		H	H	4	0.0085	0.4772	3.9677	1.07	0.3050
7		V	V	3	0.0072	0.4700	2.8570	0.90	0.3453

- **Forward Selection Procedure**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Number of Observations Read	71
Number of Observations Used	71

Summary of Forward Selection								
Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	A	A	1	0.1844	0.1844	34.3461	15.60	0.0002
2	Y	Y	2	0.1118	0.2962	22.4516	10.80	0.0016
3	G	G	3	0.1738	0.4700	2.8570	21.97	<.0001

- **Stepwise Regression Procedure**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Number of Observations Read	71
Number of Observations Used	71

Stepwise Selection: Step 1

Statistics for Entry DF = 1,69				
Variable	Tolerance	Model R-Square	F Value	Pr > F
Y	1.000000	0.0068	0.48	0.4929
J	1.000000	0.0854	6.45	0.0134
A	1.000000	0.1844	15.60	0.0002
S	1.000000	0.0786	5.88	0.0179
D	1.000000	0.0733	5.46	0.0223
F	1.000000	0.0980	7.49	0.0079
G	1.000000	0.1794	15.09	0.0002
H	1.000000	0.0145	1.01	0.3176
V	1.000000	0.0012	0.08	0.7771

Variable A Entered: R-Square = 0.1844 and C(p) = 34.3461

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4386.50023	4386.50023	15.60	0.0002
Error	69	19399	281.14269		
Corrected Total	70	23785			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	101.95006	18.32095	8705.74147	30.97	<.0001
A	-0.06354	0.01609	4386.50023	15.60	0.0002

Bounds on condition number: 1, 1

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Stepwise Selection: Step 2

Statistics for Entry DF = 1,68				
Variable	Tolerance	Model R-Square	F Value	Pr > F
Y	0.762509	0.2962	10.80	0.0016
J	0.732803	0.1912	0.57	0.4541
S	0.920145	0.2119	2.37	0.1285
D	0.995246	0.2429	5.25	0.0250
F	0.914774	0.2229	3.37	0.0709
G	0.941508	0.2930	10.44	0.0019
H	0.855899	0.1866	0.18	0.6743
V	0.705248	0.2405	5.03	0.0282

Variable Y Entered: R-Square = 0.2962 and C(p) = 22.4516

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7046.05938	3523.02969	14.31	<.0001
Error	68	16739	246.16598		
Corrected Total	70	23785			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	111.37442	17.38158	10107	41.06	<.0001
Y	0.36741	0.11178	2659.55915	10.80	0.0016
A	-0.09116	0.01724	6883.33413	27.96	<.0001

Bounds on condition number: 1.3115, 5.2458

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Stepwise Selection: Step 3

Statistics for Removal DF = 1,68				
Variable	Partial R-Square	Model R-Square	F Value	Pr > F
Y	0.1118	0.1844	10.80	0.0016
A	0.2894	0.0068	27.96	<.0001

Statistics for Entry DF = 1,67				
Variable	Tolerance	Model R-Square	F Value	Pr > F
J	0.729263	0.2997	0.33	0.5657
S	0.793608	0.3935	10.75	0.0017
D	0.900439	0.3175	2.08	0.1535
F	0.901424	0.3209	2.43	0.1238
G	0.892448	0.4700	21.97	<.0001
H	0.631758	0.3174	2.08	0.1543
V	0.610492	0.4455	18.04	<.0001

Variable G Entered: R-Square = 0.4700 and C(p) = 2.8570

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	11180	3726.51142	19.81	<.0001
Error	67	12606	188.14644		
Corrected Total	70	23785			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	79.15119	16.67857	4237.33953	22.52	<.0001
Y	0.47480	0.10037	4210.12135	22.38	<.0001
A	-0.08344	0.01516	5698.71687	30.29	<.0001
G	0.13212	0.02819	4133.47489	21.97	<.0001

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Stepwise Selection: Step 3

Bounds on condition number: 1.3836, 11.494

Stepwise Selection: Step 4

Statistics for Removal DF = 1,67				
Variable	Partial R-Square	Model R-Square	F Value	Pr > F
Y	0.1770	0.2930	22.38	<.0001
A	0.2396	0.2304	30.29	<.0001
G	0.1738	0.2962	21.97	<.0001

Statistics for Entry DF = 1,66				
Variable	Tolerance	Model R-Square	F Value	Pr > F
J	0.681983	0.4724	0.30	0.5876
S	0.486784	0.4745	0.57	0.4534
D	0.768430	0.4702	0.03	0.8669
F	0.628579	0.4775	0.95	0.3331
H	0.508819	0.4719	0.23	0.6335
V	0.212902	0.4772	0.90	0.3453

All variables left in the model are significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	A		A	1	0.1844	0.1844	34.3461	15.60	0.0002
2	Y		Y	2	0.1118	0.2962	22.4516	10.80	0.0016
3	G		G	3	0.1738	0.4700	2.8570	21.97	<.0001

- **Stepwise Regression Procedure (Fixed Y, A, and G)**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Number of Observations Read	71
Number of Observations Used	71

Stepwise Selection: Step 0

First 3 Vars Entered: R-Square = 0.4700 and C(p) = 6.0593

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	11180	3726.51142	19.81	<.0001
Error	67	12606	188.14644		
Corrected Total	70	23785			

	Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
	Intercept	79.15119	16.67857	4237.33953	22.52	<.0001
*	Y	0.47480	0.10037	4210.12135	22.38	<.0001
*	A	-0.08344	0.01516	5698.71687	30.29	<.0001
*	G	0.13212	0.02819	4133.47489	21.97	<.0001
* Forced into the model by the INCLUDE= option						

Bounds on condition number: 1.3836, 11.494

Stepwise Selection: Step 1

Variable AG Entered: R-Square = 0.5179 and C(p) = 1.8135

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	12320	3079.90448	17.73	<.0001
Error	66	11466	173.72315		
Corrected Total	70	23785			

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Stepwise Selection: Step 1

	Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
	Intercept	-32.64905	46.49152	85.67444	0.49	0.4850
*	Y	0.40240	0.10050	2784.99611	16.03	0.0002
*	A	0.01429	0.04083	21.26615	0.12	0.7275
*	G	0.96895	0.32778	1518.05299	8.74	0.0043
	AG	-0.00071071	0.00027743	1140.08367	6.56	0.0127
* Forced into the model by the INCLUDE= option						

Bounds on condition number: 164.11, 1329.4

All variables left in the model are required or significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	AG			4	0.0479	0.5179	1.8135	6.56	0.0127

B.2.2. Dummy Variables

- **Pablo**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=0 Z=0

Number of Observations Read	31
Number of Observations Used	31

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	290.67379	96.89126	0.50	0.6876
Error	27	5265.54997	195.02037		
Corrected Total	30	5556.22375			

Root MSE	13.96497	R-Square	0.0523
Dependent Mean	46.61325	Adj R-Sq	-0.0530
Coeff Var	29.95923		

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=0 Z=0

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	12.31177	29.45851	0.42	0.6793
Y	Y	1	-0.19119	0.27895	-0.69	0.4989
A	A	1	0.04116	0.04276	0.96	0.3443
G	G	1	0.01237	0.09898	0.12	0.9015
W	W	0	0	.	.	.
Z	Z	0	0	.	.	.
YW		0	0	.	.	.
AW		0	0	.	.	.
GW		0	0	.	.	.
YZ		0	0	.	.	.
AZ		0	0	.	.	.
GZ		0	0	.	.	.

- **Beijing**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=0 Z=1

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	306.45312	102.15104	1.85	0.1703
Error	20	1103.00840	55.15042		
Corrected Total	23	1409.46152			

Root MSE	7.42633	R-Square	0.2174
Dependent Mean	12.81822	Adj R-Sq	0.1000
Coeff Var	57.93575		

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=0 Z=1

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	B	33.96437	18.46754	1.84	0.0808
Y	Y	B	0.19197	0.23311	0.82	0.4199
A	A	B	-0.03050	0.01957	-1.56	0.1348
G	G	B	0.04677	0.02763	1.69	0.1060
W	W	0	0	.	.	.
Z	Z	0	0	.	.	.
YW		0	0	.	.	.
AW		0	0	.	.	.
GW		0	0	.	.	.
YZ		0	0	.	.	.
AZ		0	0	.	.	.
GZ		0	0	.	.	.

- **Peru**

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=1 Z=0

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	99.67180	33.22393	0.93	0.4574
Error	12	430.01806	35.83484		
Corrected Total	15	529.68986			

Root MSE	5.98622	R-Square	0.1882
Dependent Mean	23.63184	Adj R-Sq	-0.0148
Coeff Var	25.33117		

model selection

The REG Procedure
Model: MODEL1
Dependent Variable: R R

W=1 Z=0

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	B	-25.36947	31.14359	-0.81	0.4312
Y	Y	B	0.06830	0.11651	0.59	0.5686
A	A	B	0.03102	0.02441	1.27	0.2278
G	G	B	0.10038	0.10523	0.95	0.3589
W	W	0	0	.	.	.
Z	Z	0	0	.	.	.
YW		0	0	.	.	.
AW		0	0	.	.	.
GW		0	0	.	.	.
YZ		0	0	.	.	.
AZ		0	0	.	.	.
GZ		0	0	.	.	.

B.2.3. Comparison of Regression Models

- **All Regression Models**

all

The REG Procedure
Model: MODEL1

Test PARALLEL Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	6	67.57368	0.59	0.7398
Denominator	59	115.23011		

all

The REG Procedure
Model: MODEL1

Test COINCIDENT Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	8	725.90438	6.30	<.0001
Denominator	59	115.23011		

- **Peru - Pablo**

peru-pablo

The REG Procedure
Model: MODEL1

Test PARALLEL Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	36.23218	0.25	0.8622
Denominator	39	146.04021		

peru-pablo

The REG Procedure
Model: MODEL1

Test COINCIDENT Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	4	180.15038	1.23	0.3125
Denominator	39	146.04021		

- **Beijing – Pablo**

beijing-pablo

The REG Procedure
Model: MODEL1

Test PARALLEL Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	106.42578	0.79	0.5081
Denominator	47	135.50124		

beijing-pablo

The REG Procedure
Model: MODEL1

Test COINCIDENT Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	4	1115.13345	8.23	<.0001
Denominator	47	135.50124		

- **Peru – Beijing**

peru-beijing

The REG Procedure
Model: MODEL1

Test PARALLEL Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	55.69560	1.16	0.3392
Denominator	32	47.90708		

peru-beijing

The REG Procedure
Model: MODEL1

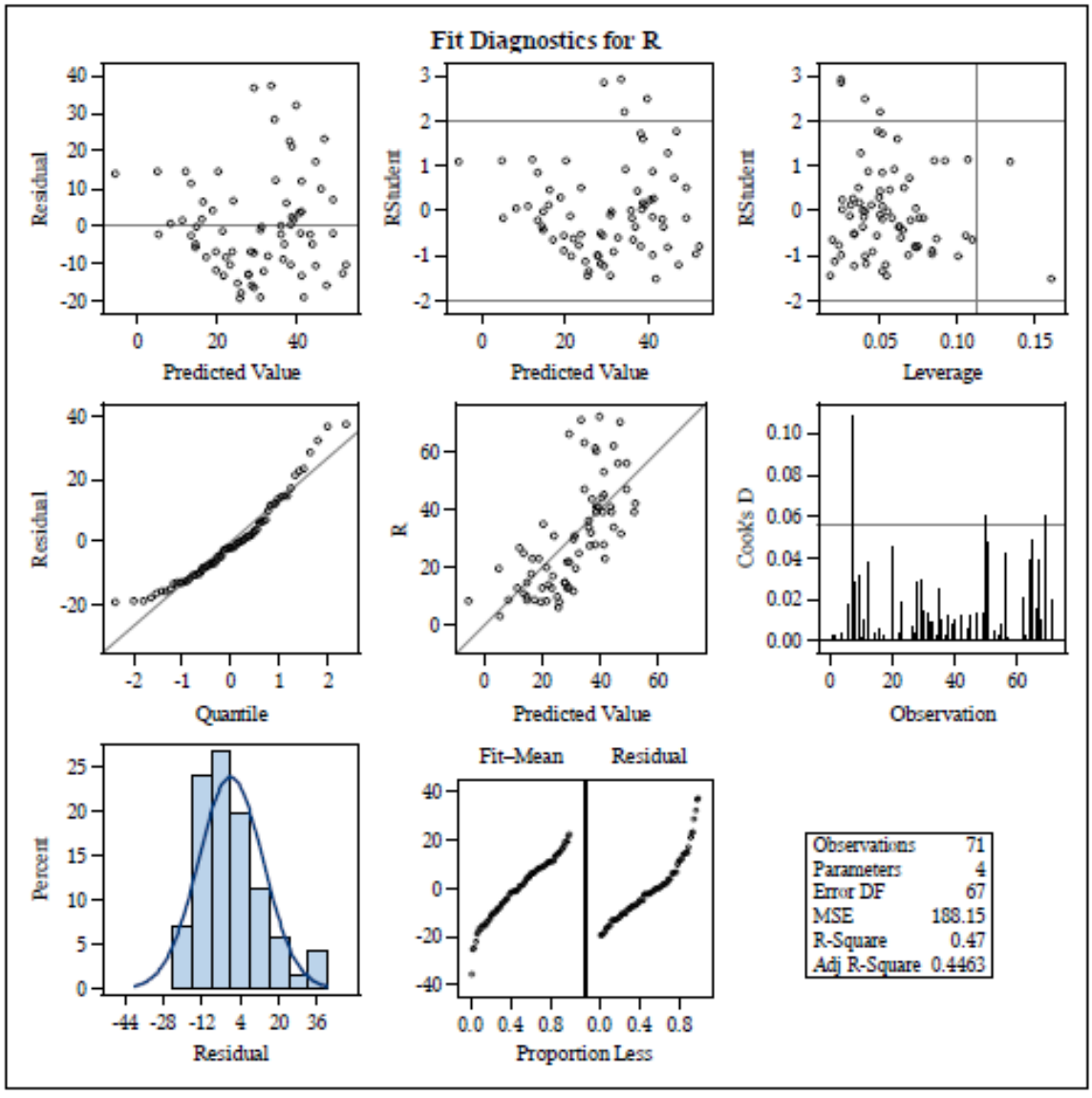
Test COINCIDENT Results for Dependent Variable R				
Source	DF	Mean Square	F Value	Pr > F
Numerator	4	86.06281	1.80	0.1539
Denominator	32	47.90708		

B.2.4. Model Diagnostics

- **Model Assumptions**

Diagnostics

The REG Procedure
Model: MODEL1



Diagnostics

The UNIVARIATE Procedure
Variable: residual (Residual)

Moments			
N	71	Sum Weights	71
Mean	0	Sum Observations	0
Std Deviation	13.4195015	Variance	180.083021
Skewness	0.91730097	Kurtosis	0.5469567
Uncorrected SS	12605.8115	Corrected SS	12605.8115
Coeff Variation	.	Std Error Mean	1.59260183

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	13.41950
Median	-2.09528	Variance	180.08302
Mode	.	Range	56.86265
		Interquartile Range	17.16037

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	-5.5	Pr >= |M|	0.2351
Signed Rank	S	-142	Pr >= |S|	0.4197

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.935018	Pr < W	0.0012
Kolmogorov-Smirnov	D	0.108073	Pr > D	0.0394
Cramer-von Mises	W-Sq	0.186465	Pr > W-Sq	0.0079
Anderson-Darling	A-Sq	1.202279	Pr > A-Sq	<0.0050

- **Outliners**

Diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Output Statistics									
						DFBETAS			
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	Y	A	G
1	6.8370	0.5048	0.0360	1.0848	0.0975	0.0008	0.0343	0.0034	-0.0483
2	-6.9271	-0.5109	0.0338	1.0820	-0.0956	-0.0426	-0.0021	0.0214	0.0704
3	-1.5197	-0.1117	0.0304	1.0945	-0.0198	-0.0035	-0.0003	-0.0007	0.0134
4	6.3941	0.4771	0.0564	1.1100	0.1166	-0.0416	0.0237	0.0480	-0.0482
5	1.8186	0.1346	0.0445	1.1102	0.0290	0.0003	-0.0013	0.0061	-0.0207
6	-12.9307	-0.9763	0.0683	1.0764	-0.2644	-0.1114	-0.1196	0.1022	0.1703
7	-18.9145	-1.5195	0.1604	1.1023	-0.6642	-0.1154	-0.5712	0.2639	0.1324
8	-13.0825	-1.0058	0.1007	1.1112	-0.3366	0.0673	-0.2336	-0.0109	0.0694
9	14.4909	1.1108	0.0923	1.0865	0.3543	-0.0674	0.1763	0.0487	-0.1476
10	4.0154	0.2983	0.0499	1.1117	0.0683	-0.0014	0.0148	0.0087	-0.0429
11	11.4365	0.8544	0.0515	1.0715	0.1992	-0.0263	-0.0129	0.0680	-0.1324
12	14.6222	1.1301	0.1065	1.1008	0.3901	-0.1854	0.1109	0.1827	-0.1297
13	-1.0826	-0.0804	0.0505	1.1181	-0.0185	-0.0014	-0.0124	0.0036	0.0055
14	-7.3270	-0.5445	0.0478	1.0955	-0.1221	0.0072	-0.0777	0.0071	0.0311
15	-0.3564	-0.0265	0.0555	1.1243	-0.0064	0.0000	-0.0045	0.0009	0.0015
16	-7.9889	-0.5988	0.0632	1.1093	-0.1555	-0.0206	-0.1125	0.0425	0.0457
17	-4.7764	-0.3573	0.0625	1.1241	-0.0923	-0.0394	0.0510	0.0011	0.0721
18	-2.6262	-0.1971	0.0696	1.1387	-0.0539	-0.0192	0.0390	-0.0046	0.0352
19	0.6570	0.0494	0.0735	1.1461	0.0139	0.0015	-0.0106	0.0047	-0.0087
20	13.7971	1.0821	0.1337	1.1426	0.4251	-0.1377	-0.3109	0.2978	-0.2019
21	1.5656	0.1164	0.0519	1.1192	0.0272	0.0015	-0.0173	0.0097	-0.0179
22	-5.8378	-0.4375	0.0649	1.1226	-0.1153	-0.0494	0.0714	0.0001	0.0846
23	-11.7374	-0.8926	0.0837	1.1047	-0.2698	-0.1764	0.1509	0.0694	0.1838
24	-2.1958	-0.1652	0.0751	1.1463	-0.0471	0.0044	0.0339	-0.0239	0.0277
25	-0.1235	-0.009121	0.0404	1.1067	-0.0019	0.0009	0.0010	-0.0014	0.0002
26	-15.2097	-1.1229	0.0211	1.0058	-0.1649	-0.0456	0.0856	-0.0093	0.0521
27	-10.4681	-0.7700	0.0237	1.0496	-0.1200	-0.0174	0.0746	-0.0247	0.0322
28	14.6079	1.1152	0.0848	1.0769	0.3394	-0.2107	-0.1865	0.2947	-0.0567
29	-19.1252	-1.4453	0.0541	0.9912	-0.3458	0.1566	0.1246	-0.1637	-0.2282

Diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Output Statistics									
						DFBETAS			
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	Y	A	G
30	-15.8523	-1.1837	0.0410	1.0182	-0.2448	0.1156	0.0933	-0.1288	-0.1391
31	-16.4815	-1.2268	0.0335	1.0040	-0.2282	0.0798	0.0942	-0.0994	-0.1174
32	-19.4024	-1.4388	0.0180	0.9558	-0.1946	0.0186	0.0849	-0.0646	0.0033
33	-8.2669	-0.6277	0.0863	1.1350	-0.1929	0.1393	0.0763	-0.1472	-0.1037
34	6.0712	0.4515	0.0503	1.1045	0.1039	0.0503	-0.0438	-0.0331	0.0221
35	-17.9739	-1.3538	0.0515	1.0035	-0.3156	0.0207	0.2309	-0.0997	-0.0768
36	-13.3480	-0.9935	0.0408	1.0433	-0.2048	0.0931	0.1254	-0.1345	-0.0520
37	-6.7961	-0.5011	0.0332	1.0819	-0.0928	0.0604	-0.0005	-0.0540	-0.0484
38	-8.3549	-0.6427	0.1097	1.1635	-0.2256	0.1852	0.0780	-0.1927	-0.1037
39	-7.0012	-0.5368	0.1055	1.1668	-0.1844	0.1380	0.0767	-0.1469	-0.0952
40	-12.1934	-0.9086	0.0453	1.0584	-0.1979	0.1314	-0.0199	-0.1030	-0.1329
41	0.4566	0.0335	0.0255	1.0898	0.0054	0.0036	0.0013	-0.0036	-0.0000
42	-10.3018	-0.7784	0.0744	1.1062	-0.2208	-0.1605	-0.0549	0.1773	-0.0533
43	-2.0953	-0.1562	0.0576	1.1253	-0.0386	-0.0283	-0.0165	0.0325	-0.0050
44	-13.1869	-0.9734	0.0253	1.0292	-0.1569	-0.0468	-0.0609	0.0702	-0.0687
45	-10.3234	-0.7793	0.0727	1.1040	-0.2181	-0.2010	0.0230	0.1620	0.0765
46	-2.1362	-0.1610	0.0782	1.1503	-0.0469	-0.0404	0.0144	0.0293	0.0144
47	-10.7721	-0.8136	0.0730	1.1008	-0.2283	-0.2075	-0.0215	0.1919	0.0360
48	0.0583	0.004302	0.0394	1.1056	0.0009	0.0007	-0.0001	-0.0006	-0.0002
49	12.2889	0.9227	0.0592	1.0724	0.2315	0.1740	-0.0946	-0.1163	-0.0432
50	28.5521	2.1958	0.0501	0.8433	0.5044	0.3851	-0.1834	-0.2627	-0.1054
51	36.7372	2.8531	0.0249	0.6840	0.4558	0.2986	-0.1015	-0.1786	-0.2062
52	3.3983	0.2492	0.0256	1.0859	0.0404	0.0210	0.0109	-0.0235	0.0103
53	6.9128	0.5184	0.0653	1.1179	0.1370	0.1105	0.0460	-0.1188	0.0069
54	-4.8268	-0.3610	0.0622	1.1236	-0.0930	-0.0741	0.0079	0.0658	-0.0036
55	11.8630	0.8824	0.0425	1.0583	0.1859	0.1316	-0.0172	-0.1161	0.0207
56	21.1545	1.6109	0.0616	0.9700	0.4129	-0.1606	-0.0317	0.1127	0.3437
57	-4.8665	-0.3595	0.0388	1.0962	-0.0722	0.0154	0.0114	-0.0120	-0.0521
58	1.7179	0.1263	0.0314	1.0953	0.0227	-0.0048	0.0054	0.0007	0.0168

Diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: R R

Output Statistics									
						DFBETAS			
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	Y	A	G
59	3.7016	0.2725	0.0326	1.0929	0.0500	-0.0072	0.0237	-0.0053	0.0339
60	-2.3758	-0.1766	0.0519	1.1180	-0.0413	0.0066	-0.0034	-0.0000	-0.0343
61	-1.9986	-0.1476	0.0398	1.1046	-0.0300	0.0091	-0.0149	-0.0007	-0.0215
62	-12.6053	-0.9596	0.0839	1.0968	-0.2904	0.0506	-0.2093	0.0560	-0.2058
63	-8.8879	-0.6517	0.0199	1.0561	-0.0928	-0.0041	-0.0454	0.0188	-0.0308
64	22.7148	1.7254	0.0516	0.9387	0.4026	-0.1694	0.0080	0.1111	0.3328
65	37.4602	2.9164	0.0249	0.6706	0.4658	-0.1589	-0.0014	0.1264	0.2904
66	17.1724	1.2822	0.0375	0.9999	0.2530	0.0221	0.1574	-0.0921	0.1442
67	23.2270	1.7632	0.0486	0.9286	0.3986	0.2817	0.0970	-0.3044	0.0813
68	9.8221	0.7396	0.0691	1.1037	0.2014	-0.0651	0.1305	-0.0045	0.1460
69	32.3126	2.4969	0.0404	0.7713	0.5122	-0.1735	0.3007	0.0249	0.3189
70	2.3396	0.1725	0.0367	1.1005	0.0337	-0.0128	0.0144	0.0044	0.0237
71	-15.9266	-1.1973	0.0534	1.0295	-0.2843	0.0062	-0.2098	0.0885	-0.1668

- **Collinearity**

Diagnostics

The REG Procedure

Number of Observations Read	71
Number of Observations Used	71

Correlation					
Variable	Label	Y	A	G	R
Y	Y	1.0000	0.4873	-0.3113	0.0827
A	A	0.4873	1.0000	-0.2419	-0.4294
G	G	-0.3113	-0.2419	1.0000	0.4236
R	R	0.0827	-0.4294	0.4236	1.0000

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	79.15119	16.67857	4.75	<.0001	0
Y	Y	1	0.47480	0.10037	4.73	<.0001	1.38356
A	A	1	-0.08344	0.01516	-5.50	<.0001	1.32713
G	G	1	0.13212	0.02819	4.69	<.0001	1.12051

- **Y, A, and G Statistics**

Diagnostics

The UNIVARIATE Procedure
Variable: Y (Y)

Moments			
N	71	Sum Weights	71
Mean	59.4365742	Sum Observations	4219.99677
Std Deviation	19.2125728	Variance	369.122955
Skewness	0.73095498	Kurtosis	0.00451454
Uncorrected SS	276660.758	Corrected SS	25838.6068
Coeff Variation	32.3244956	Std Error Mean	2.2801129

Basic Statistical Measures			
Location		Variability	
Mean	59.43657	Std Deviation	19.21257
Median	55.00000	Variance	369.12295
Mode	48.35526	Range	82.84891
		Interquartile Range	25.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	26.06738	Pr > |t|	<.0001
Sign	M	35.5	Pr >= |M|	<.0001
Signed Rank	S	1278	Pr >= |S|	<.0001

Diagnostics

The UNIVARIATE Procedure
Variable: A (A)

Moments			
N	71	Sum Weights	71
Mean	1132.1453	Sum Observations	80382.3166
Std Deviation	124.579185	Variance	15519.9733
Skewness	-0.2644905	Kurtosis	-0.377717
Uncorrected SS	92090860.5	Corrected SS	1086398.13
Coeff Variation	11.0038159	Std Error Mean	14.7848292

Basic Statistical Measures			
Location		Variability	
Mean	1132.145	Std Deviation	124.57918
Median	1151.259	Variance	15520
Mode	1183.224	Range	514.00000
		Interquartile Range	139.97512

Note: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	76.5748	Pr > |t|	<.0001
Sign	M	35.5	Pr >= |M|	<.0001
Signed Rank	S	1278	Pr >= |S|	<.0001

Diagnostics

The UNIVARIATE Procedure
Variable: G (G)

Moments			
N	71	Sum Weights	71
Mean	129.429061	Sum Observations	9189.4633
Std Deviation	61.5681394	Variance	3790.63579
Skewness	-0.1652438	Kurtosis	-1.3734424
Uncorrected SS	1454728.11	Corrected SS	265344.505
Coeff Variation	47.5690228	Std Error Mean	7.30679385

Basic Statistical Measures			
Location		Variability	
Mean	129.4291	Std Deviation	61.56814
Median	137.0000	Variance	3791
Mode	63.1579	Range	210.12935
		Interquartile Range	119.87104

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	17.71352	Pr > |t|	<.0001
Sign	M	35.5	Pr >= |M|	<.0001
Signed Rank	S	1278	Pr >= |S|	<.0001