How Red-Light Camera Studies are Biased by Regression to the Mean

Dale Gedcke

Red-Light Camera studies are replete with concerns about the bias introduced by “regression to the mean”. For those not steeped in statistics, that term sounds awfully obscure and meaningless. Nevertheless, regression to the mean is an important effect that can lead to erroneously optimistic conclusions when simply comparing accident rates before and after camera installation. This article offers a simple explanation of the effect, why it is important, and what some investigators have done to circumvent the bias. For further discussions on the issue, one can consult references 1 and 2.

Accidents at intersections controlled by traffic lights are rare. The accident rate is typically in the range of 0.2 to 3 accidents for every million vehicles transiting the intersection. As a result, accidents occur randomly, and are subject to random fluctuations around the average accident rate for any selected intersection, or group of intersections. If you check the monthly accident rates at a particular intersection, you will find that number fluctuates up and down around some average. Statisticians prefer to call that average the “mean”. Small deviations from the mean are much more prevalent than large excursions.

These random fluctuations around the mean contribute to the bias in Red-Light Camera studies, because of the scheme by which intersections are selected for Red-Light Camera treatment. A municipality concerned about safety, will naturally select the intersections with the highest accident rates for camera installation. If intersection A had the highest accident rate out of 50 similar intersections in 2006, will it also have the highest accident rate in 2007 (if cameras were not installed)? The answer is: most likely, intersection A will have a lower accident rate in 2007 than in 2006. By selecting the intersection with the highest accident rate in 2006, one has chosen a case where an extremely large random deviation from the mean was very likely involved. That large positive deviation is unlikely to happen again in 2007. Thus a comparison between 2006 and 2007 will indicate a decrease in the accident rate, even if a Red-Light Camera was not installed at the end of 2006. This is the “regression to the mean” effect. Because an intersection with an abnormally large positive excursion from the mean was selected in 2006, the accident rate in 2007 will tend to decrease and come closer to the mean.

A practical example will clarify the concept. Rainfall is subject to random variations and is randomly distributed over the area of a city. Table 1 shows the measured rainfall in Knoxville, Tennessee for the month of July 2006 for the six measurement sites throughout the city. (More complete data can be found in reference 3.)

Table 1. Knoxville July Rainfall in Inches: 2006

	Acker Place	KAT Station	Lowes Creek	Walden Drive	Williams Creek	NOAA Airport	Area Average
Jul-06	4.06	3.09	3.31	4.89	3.35	3.94	3.773

Let’s suppose that you have developed an ultrasonic device that you think will reduce rainfall. To test the device, you select the rain gauge site with the highest rainfall (Walden Drive) and install your device near that gauge at the end of June 2007. At the end of July 2007 you check the recorded rainfall and compare that result to 2006. The result is in Table 2.

Eureka! The ultrasonic device has made a dramatic difference at the Walden Drive rain gauge. The rainfall at that location has dropped by 37.2%, while the average across all six gauges has decreased only 5.4%. You are tempted to ascribe the 37.2% - 5.4% = 31.8% net decrease to your fantastic invention. But, you and I both know that your invention was pure fiction. You did not install any ultrasonic device at the Walden Drive location. So, why did the rainfall measured at Walden Drive decrease by a phenomenal 37.2%, while the average for all six sites declined by only 5.4%? The answer is: regression to the mean. By selecting the gauge that had the largest positive random deviation from the mean, you virtually guaranteed that it would fall closer to the mean on the next measurement*.

Table 2. Knoxville July Rainfall in Inches: 2007 vs. 2006

	Acker Place	KAT Station	Lowes Creek	Walden Drive	Williams Creek	NOAA Airport	Area Average
Jul-06	4.06	3.09	3.31	4.89	3.35	3.94	3.773
Jul-07	3.34	4.15	2.64	3.07	3.5	4.71	3.568
% Change	-17.7	34.3	-20.2	-37.2	4.5	19.5	-5.4

Now, relate this example to what happens with Red-Light Camera installations. The sites that receive the cameras are typically selected because they had the highest accident rates in the previous year. That is similar to selecting the rain gauge with the highest recorded rainfall. Those camera sites will tend to show a lower accident rate in the year following camera installation as a result of regression to the mean, regardless of the camera installation.

How does one separate the true effect of camera installation from the decrease caused by regression to the mean? Garber *et al.*[2] recommend the Empirical Bayes method of analysis[4], because it tends to circumvent regression to the mean. For the six Virginia jurisdictions studied by Garber *et al.*, the average decrease in the angle collision rate was 42%, when using a simple before/after comparison. But, the Empirical Bayes analysis yielded a much smaller, 8% decrease. The authors claim that the Empirical Bayes analysis was successful in eliminating the bias caused by regression to the mean.

The moral of the story is: “Beware of regression to the mean!” When accident rates are reported as simple before/after comparisons for Red-Light Camera intersections, a substantial portion of the change in accident rates can be due to regression to the mean. This effect tends to make the cameras look more effective than they actually were.

References

1. Lawrence E. Decina; Libby Thomas; Raghavan Srinivasan, Ph.D.; and Loren Staplin, Ph.D., *Automated Enforcement: A Compendium of Worldwide Evaluation of Results*, Report No. DOT HS 810 763, Sept. 2007, posted at www.nhtsa.gov.

2. Nicholas J. Garber, John S. Miller, R. Elizabeth Abel, Saeed Eslambolchi, and Santhosh K. Korukonda, *The Impact of Red Light Cameras (Photo-Red Enforcement) on Crashes in Virginia*, Virginia Transportation Research Council, Charlottesville, VA, June 2007.

3. Rainfall Data, City of Knoxville Engineering Division, Stormwater Engineering Section, posted at http://www.cityofknoxville.org/engineering/stormwater/rainfall/default.asp

4. Peter M. Lee, *Bayesian Statistics: An Introduction*, Hodder Arnold, London England, 2004.

* NOTE: Regression to the mean works from both extremes. If the lowest value is chosen, it is likely to increase in the next measurement, as it moves closer to the mean. More generally, the average of the measurements approaches the true population mean, as more measurements are included in the average.