# Chapter 1. What is what

## The safety property of an intersection.

The intersection of Eglinton and Don Mills roads is mentioned in Section 1.1 where ‘safety’ is defined. The number of reported accidents reported at this intersection in each of 52 ‘police reporting periods’[[1]](#footnote-1) is in the ‘Toronto Intersection.xls’ spreadsheet. Based on this data what was the estimate of the safety of this intersection in PRP 35?

## The safety of Colorado two-lane rural roads

Two data sets were used in the numerical examples of Section 1.4. For the Connecticut drivers data the computations were described in detail whereas for the Colorado two-lane rural road data only the final results were given. The intent of the following problems is to check these results. The requisite data for 0-1 miles long rural two-lane road segments are in the ‘Colorado Segments.xls’ spreadsheet.

### Estimating E{μ} and σ{μ}.

Confirm that the estimates of E{μ} and σ{μ} are those given in Section 1.4.2.

### How many ‘deviant’ units are there?

‘Screening’ is the activity of identifying units that are unusually unsafe. The intent is to examine the identified units in order to decide whether an intervention is called for. When these units are road segments, intersections, grade crossings etc. they are called ‘blackspots’, ‘sites with promise’ and similar names.

What is deemed to be ‘unusually unsafe’ depends mostly on what the potential interventions are. If the cost of the intervention is proportional to segment length (e.g. pavement resurfacing) then segments with high μ per unit of length are of interest; if cost is proportional to amount of traffic (e.g. speed enforcement) then segments with high μ per unit of travel might be of interest; if the cost of the intervention does not depend on the size of the unit or the amount of traffic (e.g. change of speed limit or of the intergreen time at a signalized intersection) then units with high μ are of interest. The latter is the option pursued in this exercise.

A μ is unusually high if it is much larger than the average μ for the population; that is, if it is larger than c×E{μ} where c is some number larger than 1.The larger is the c the more unsafe are the sought units. To design a sensible screening program it helps to know how many units in the population are expected to have a μ>c×E{μ}. Because the μ of units is never known and can be only imperfectly estimated, if the unusually unsafe units are rare, just as the needle in the haystack, they will be difficult to find.

Following the example in Section 1.4.3 determine the proportion and the expected number of segments Colorado road segments with μ>c×E{μ} when c=2, 3, 4, and 5. What ‘c’ corresponds to the top 5th percentile of μ’s?

### How good is a screen

Suppose that the aim is to identify segments the μ of which is more than five times the population mean, that is, segments with μ>9.3 accidents in 13 years. As noted earlier, the μ of a unit is never known; here only k, the number of recorded accidents is known. Suppose that segments with k≥8 are selected for closer examination (‘engineering study’). The questions are how many are expected to have μ>5×E{μ} - these are the ‘correct positives’? How many segments with k≥8 are expected to have μ<5×E{μ} - these are the ‘false positives’? How many segments with μ>5×E{μ} are expected to have k<8 - these are the ‘false negatives’)?

### A discrepancy

There are two ways to determine the number of false negatives. One is to subtract from the number of segments with k≧8 and expected to have μ>9.3, as determined in problem b, the number of correct positives as determined in problem c. The other way is to compute the number of segments with μ>9.3 amongst segments with k<8 and add these up. One might expect these two ways to produce identical results but this is not so. How large is the discrepancy and what explains it?

### The expected accident reduction

Suppose that segments with k≥10 [accidents in 13 years] are to be treated and the CMF[[2]](#footnote-2) of the treatment is 0.92. How many accidents/year should one expect to save?

### A CMF estimate based on ‘Before-After’ data

Suppose that a certain treatment was implemented on all segments with k≥15 [accidents in 13 years] and that there was no change in any safety-relevant traits from the ‘before’ to the ‘after’ treatment period. During a five year ‘after’ period 193 accidents were recorded at these sites. What is the corresponding CMF estimate?

### Is the ‘Gamma assumption’ well supported?

As discussed in Section 1.4.6 of the book, if the μ’s in this population of segments were Gamma distributed, and if the accident counts on each segment obeyed the Poisson distribution then the distribution of accident counts in this population of segments would obey the Negative Binomial distribution (equation 1-6). If the Negative Binomial distribution applied, and using the parameter estimates of ‘a’ and ‘b’ obtained earlier ,what is the expected number of segments with k=0, 1, …,48? Does the correspondence between the observed and the expected number of segments for each k support the Gamma assumption? To check conduct a χ2 test (of the distribution of segment counts as a whole).

### How good is the Empirical Bayes estimator in this case?

The expected accident frequency μ of a unit (a road segment in this case) is of interest for a variety of purposes. If the unit recorded k accidents in a period what should we estimate its μ to be? The traditional answer is . This answer would be a reasonable if the unit did not change from period to period and if was the average of a large number of k’s. This approach to estimation is not reasonable in a world where units change over time and averaging over many periods is not practical. In this real world the Empirical Bayes estimator is to be preferred. When the aforementioned Poisson and Gamma assumptions hold, the EB estimator of is (see the first part of equation 1-5 in the book.) The question is how well does it work? To answer one can use the data in Table 1.

Table . Colorado Road Segments 1997/1998

|  |  |  |
| --- | --- | --- |
| 1997 | | 1998 |
| Accidents, k | Segments, n(k) | Average |
| 0 | 4251 | 0.173 |
| 1 | 710 | 0.559 |
| 2 | 207 | 1.227 |
| 3 | 69 | 1.928 |
| 4 | 37 | 2.811 |
| 5 | 16 | 3.188 |
| 6 | 13 | 3.077 |
| 7 | 4 | 5.250 |
| 8 | 5 | 3.800 |
| 9 | 3 | 9.333 |
| 10 | 2 | 5.500 |
| 11 | 2 | 8.000 |
| 12 | 2 | 6.000 |
| 14 | 1 | 13.000 |
| 25 | 1 | 15.000 |
| Grand Total | 5323 |  |

To explain, the n(k) column is the number of Colorado road segments with k accidents in 1997. In that year there were 1837 accidents on all the 5323 segments and in 1998 the total was very similar, 1850.

As shown in the table, segments which in 1997 recorded 0 accidents had on average 0.173 accidents in 1998. Segments which in 1997 had 1 accident had on average 0.559 accidents per segment in 1998 etc. This is regression-to-mean in action. It is plain that the number of accidents in one year, the k, is not a good predictor for the next year.

The question is whether the EB estimates are close to the average what has materialized. (Because the accident total in 1997 was similar to that in 1998 there is no real need for a ‘time trend’ correction.). Should the EB estimates prove reasonably good, they would correct for the regression-to-mean bias.

1. 1 PRP=4 weeks. [↑](#footnote-ref-1)
2. Crash Modification Factor or Function ≡ CMF ≡(Expected accidents with the treatment implemented)/(Expected accidents if treatment is not implemented). [↑](#footnote-ref-2)