# Chapter 4. Curve fitting

## The safety property of an intersection revisited.

Problem 1 for Chapter 1 reads as follows: “… 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’ file. Based on this data what was the estimate of the safety of this intersection in PRP 35?” The running average and Holt smoothing were used to answer. Both are curve fitting methods.

Here the same data (‘Toronto Intersection.xls’ ) is to be used and a curve fitted by the Nadaraya Watson kernel regression using bandwidth=2 and bandwidth=5. Use the first workpage in the ‘Nadaraya Watson.xls’ with ‘PRP number’ as X and the ‘# Acc.’ As Y.

## How the number of accidents depends on Segment Length and on AADT

In Chapter 3 (using the data in the ‘State data and attributes.xls’ file) plots of accidents/segment as a function of Segment Length were produced for AADT bins 1000<AADT<2000 and 2000<AADT<3000.

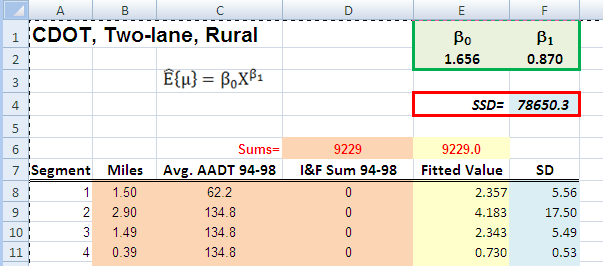
1. Produce the same plots by the 2D Nadaraya-Watson procedure.
2. Produce smoothed estimates of accidents/segment the 3D N-W procedure. Show plots against AADT and Segment Length.
3. Nadaraya-Watson

Data: Exercises for Chapter 4. ‘OLS Fit’

One way to estimate σ{μ} is by:

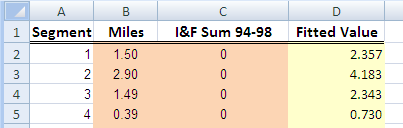
|  |  |  |
| --- | --- | --- |
|  |  |  |

In chapter 5 we will use SOLVER to obtain least-squares estimates of parameters β0 and β1. The top part of that C-F spreadsheet is in Figure 1.



Figure

I copied the Segment Length (Miles) from column B, the observed accident count (I&F Sum) from column D, and the Fitted Values from Column E into spreadsheet: Data: Exercises for Chapter 4. ‘OLS Fit’. The top part of the spreadsheet is in Figure 2



Figure

1. Compute for all segments and plot as a function of fitted values. Can you discern some orderly relationship?
2. Copy the fitted values and the estimates of σ{μ} into columns A and B of the ‘Nadaraya Watson non parametric spreadsheet’. Sort the data (both columns) in the increasing order of the Fitted Values. Click on the command button and follow instructions.
3. SOLVER and the ‘Calibration Factor’

Data: ‘Predicted and Observed’ spreadsheet.

To make use of the SPFs in the Highway Safety Manual one has to have a jurisdiction-specific ‘calibration factor, C’ defined as:

|  |  |  |
| --- | --- | --- |
|  |  |  |

The summation is over a set of sites where information is collected about specified traits specified in the HSM and used for the HSM prediction.

A State collected data at 350 segments of multilane rural roads and used these to compute the HSM predicted yearly crash frequency and ascertained the number of observed crashes in three years. The resulting data are in: the ‘Predicted and Observed’ spreadsheet.

1. Determine C.
2. Prepare a graph with ‘Predicted/year’ on the horizontal axis and ‘Observed in three years’ on the vertical axis. Does the relationship look like a straight line?
3. To check whether it is a straight line estimate the parameters β0 and β1 of the model equation

|  |  |  |
| --- | --- | --- |
|  |  |  |

If it is a straight line the β0/3 should be close to C and β1 should be close to 1.

The estimation will be by maximizing the natural logarithm of the Poisson likelihood function. As will be shown in Chapter 7, the (abridged) Poisson log-likelihood function is:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Replacing the μ’s in equation 4 by the model equation turns the log-likelihood into a function of β0 and β1, i.e. into. Fitting this model function to data amounts to finding those values of β0 and β1 which maximize .

Is the function a straight line? What are the implications of your finding for the HSM calibration procedure?

1. 1 PRP=4 weeks. [↑](#footnote-ref-1)