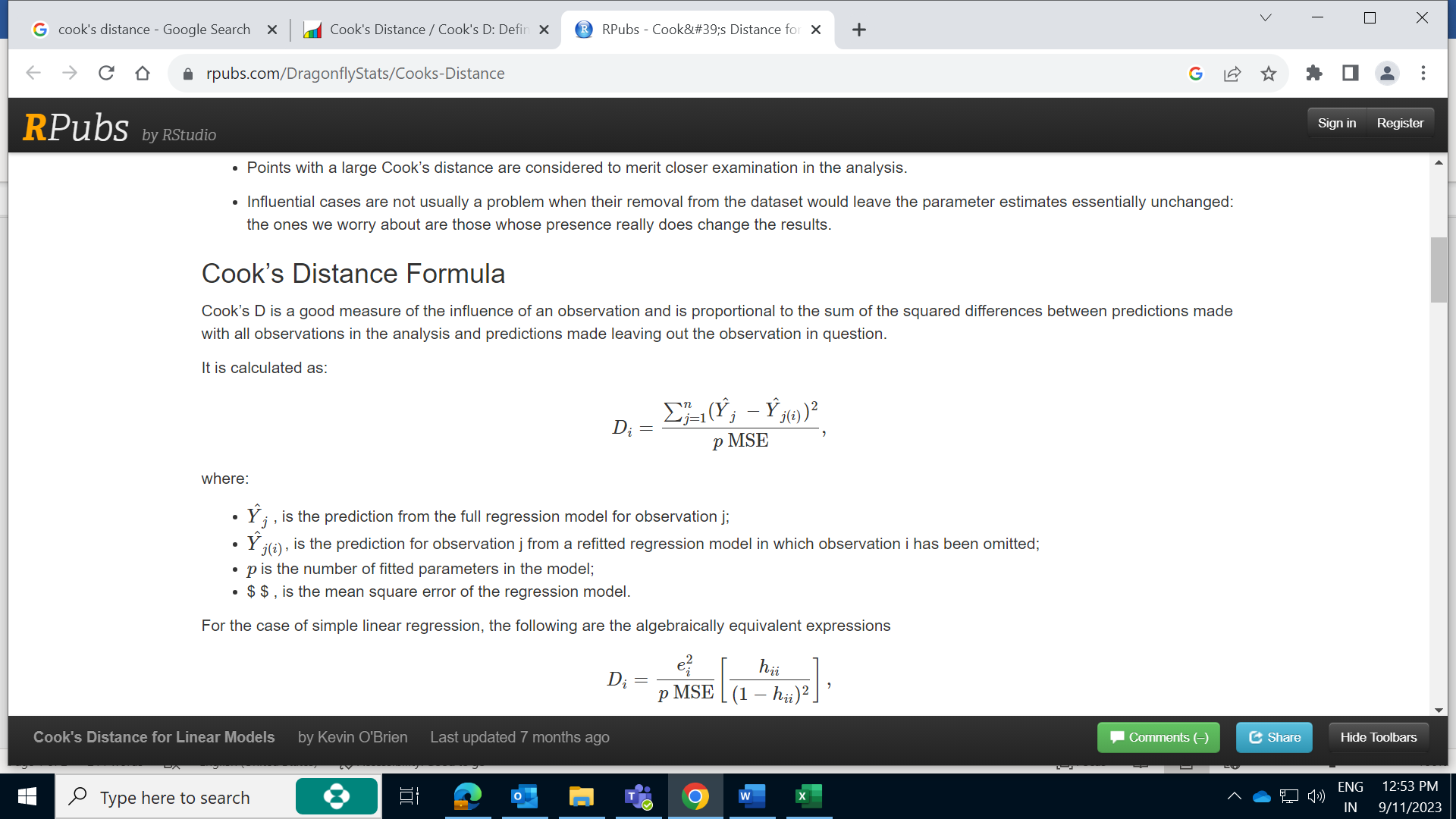
**Cook’s Distance / Cook’s D**

* Cook’s distance or Cook’s D is a commonly used estimate of the influence of a data point when performing least squares regression analysis.
* Cook’s distance is useful for identifying outliers in the X values (observations for predictor variables). It also shows the influence of each observation on the fitted response values.

Cook’s Distance Formula

Cook’s D is a good measure of the influence of an observation and is proportional to the sum of the squared differences between predictions made with all observations in the analysis and predictions made leaving out the observation in question.

It is calculated as:

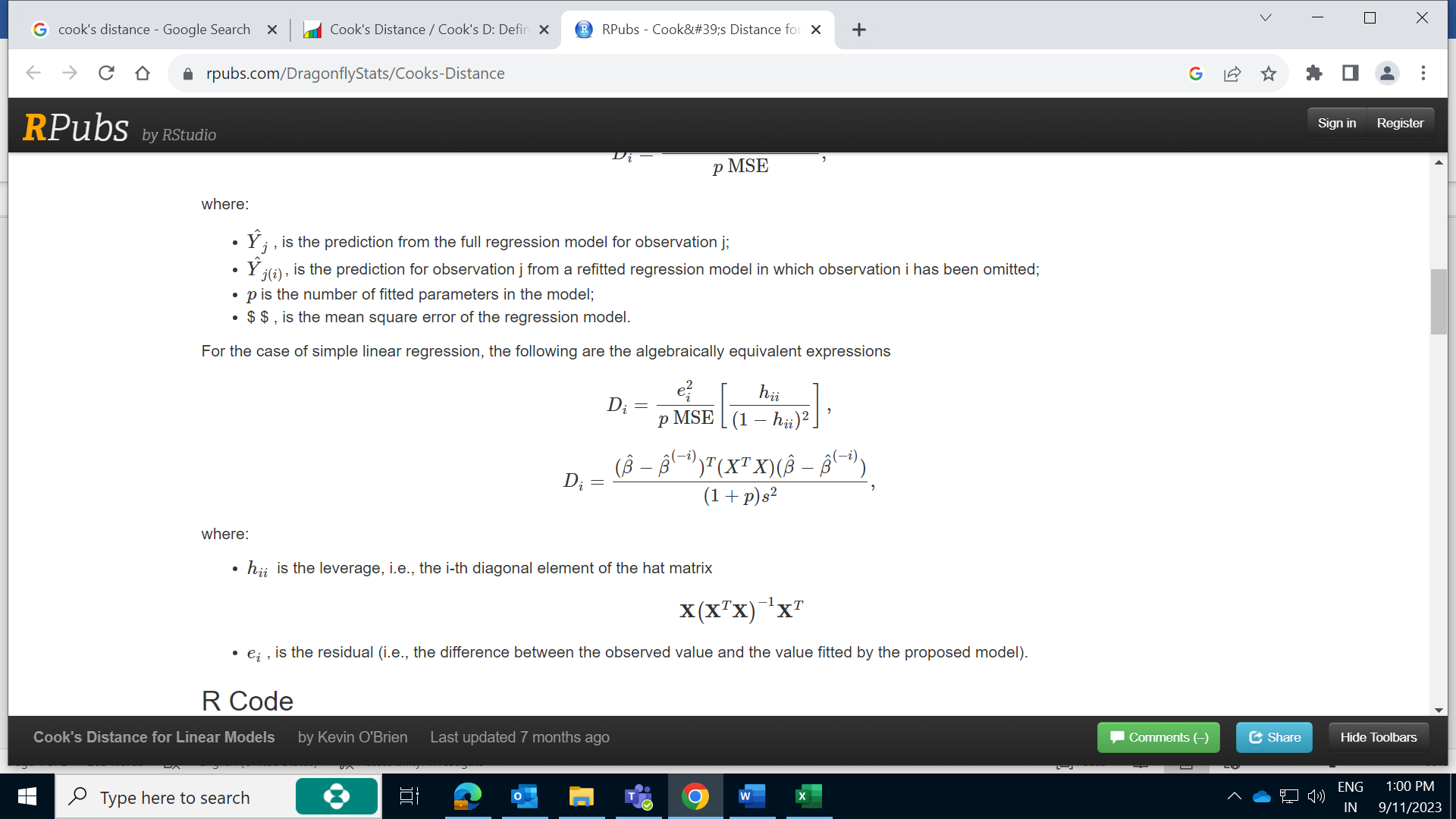


Where:

* Y^j, is the prediction from the full regression model for observation j;
* Y^j(i) is the prediction for observation j from a refitted regression model in which observation i has been omitted.
* p is the number of fitted parameters in the model;
* $ $, is the mean square error of the regression model.

For the case of simple linear regression, the following are the algebraically equivalent expressions

where:



* hii, is the leverage, i.e., the i-th diagonal element of the hat matrix

A black text with a white background

Description automatically generated

* ei, is the residual (i.e., the difference between the observed value and the value fitted by the proposed model).

**Proc GAM Model**

The MODEL statement specifies the dependent variable and the independent effects you want to use in the model.

Specify the independent parametric variables inside the parentheses of PARAM( ). The parametric variables can be either classification variables or continuous variables.

|  |  |
| --- | --- |
| **Smoothing Effect** | Description |
| SPLINE | * It is used to specify non-parametric smoothing of predictor variable. * Non-Parametric does not assume a specific mathematical form for the relationship between predictor and response. * It allows the data to determine the shape of the curve. * It uses technique like cubic spline or LOESS to estimate smooth curve. |
| LOESS | Fits a local regression with the variable and with DF=*number* |
| SPLINE2 | Fits a bivariate thin-plate smoothing spline with *variable1* and *variable2* and with DF=*number* |
| PARAM | It specifies a parametric function for a predictor value.  It assumes relationship between predictor and  Response variable follows a predefine parametric face, such as linear, quadratic or cubic |

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