International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 12 Issue 01 January 2024, Page no. 3937-3940

Index Copernicus ICV: 57.55, Impact Factor: 7.362

DOI: 10.47191/ijmcr/v12i1.02

Computational Statistics with MATLAB

Prof. Nitin Geete
Asst. Professor [HOD] Gyan Ganga College of Excellence, Jabalpur

ARTICLE INFO

Published Online:
02 January 2024

Corresponding Author:
Prof. Nitin Geete

ABSTRACT

The Definition of Computational Statistics Clearly, computational statistics has a connection to the field of statistics as a whole. Determining what we mean by the area of statistics is therefore necessary before we define computational statistics in its appropriate sense. At its most fundamental level, statistics deals with turning unprocessed data into knowledge [Wegman, 1988]. Any scientist who is faced with an application that calls for the analysis of raw data must consider issues like:

- What data should be collected to answer the questions in the analysis?
- How much information needs to be gathered?
- What conclusions can be made based on the information?
- How much of those conclusions can be believed? The science of uncertainty is a topic that statistics addresses, and it can assist the scientist in answering these queries.

KEYWORDS:

I. INTRODUCTION

Scientists are familiar with and often use many traditional statistical techniques developed over the past century, such as regression, hypothesis testing, parameter estimation, confidence intervals, etc. Enron and Tibshirani, 1991]. Now what exactly do we mean by computational statistics? Here, we once again follow Wegmann's [1988] definition. Computational statistics is, according to Wegman, a set of methods with "an emphasis on the exploitation of information science in the development of new statistical methods". After the creation of affordable computer hardware in the 1980s, many of these approaches became practical. Thanks to the computer revolution, scientists and engineers can now store and process massive amounts of data. However, most of the time there is no clear plan as to how this data will be used once it is received for research. For example, when we do data analysis today, we often collect data before designing an investigation to extract useful information. The traditional approach, in contrast, has been to first design the study based on the research objectives and then collect the necessary data. The data sets that analysts must deal with today are often highly dimensional and very large, due to the low cost of storage and acquisition. Many traditional statistical tools fall short in such cases. Wegman [1988] lists non-parametric functional estimation, parallel coordinates for high-dimensional data representation, and data sets as examples of computational statistical techniques.

1.2 Traditional Statistics Vs Computational Statistics

Traditional Statistics	Computational Statistics
Small to moderate sample size	Large to very large sample size
Independent, identically distributed data sets	No homogeneous data sets
One or low dimensional	High dimensional
Manually computational	Computationally intensive
Mathematically tractable	Numerically tractable
Well focused questions	Imprecise questions
Strong unverifiable assumptions: Relationships (linearity, additivity) Error structures (normality)	Weak or no assumptions: Relationships (nonlinearity) Error structures (distribution free)
Statistical inference	Structural inference
Predominantly closed form algorithms	Iterative algorithms possible
Statistical optimality	Statistical robustness

Predominantly closed form algorithms	Iterative algorithms possible

1.3 MATLAB Computational Statistical Tools

Most of the algorithms in this book's coverage are not compatible with MATLAB. As a result, we offer functions that fulfill most of the instructions specified in the text. Note that these functions are slightly different from the MATLAB code provided in the examples. Functions often allow the user to implement general case algorithms. A list of features and the intended use of each is provided in Appendix F. At the end of the chapter, we also provide a summary of relevant works.

1.4 What Is MATLAB, exactly?

Math Works, Inc. Created the MATLAB technical computing environment to calculate and display data. A basic data element of this interactive system and programming language is an array, which can be a scalar, vector, matrix, or multidimensional array. It provides programming features comparable to other computer languages, as well as basic matrix operations (e.g., functions, control flow, etc.). To help the reader understand the algorithms in the text, we have included a quick overview of MATLAB in this Appendix. We do not claim that this introduction is exhaustive, and we encourage the reader to look elsewhere for additional information about MATLAB. The MATLAB documentation is top notch and the tutorials should be beneficial to the reader. For a full description, we also suggest the work of Hanselman and Littlefield in MATLAB (199, 1998, 2001). Marchand [1999] should be used if the reader needs to learn more about the GUI and graphical features of MATLAB.

MATLAB will run on Windows, Unix and Linux platforms. Although we're focusing on the Windows version in this article, much of the wisdom is universal. The standard MATLAB software package includes many data analysis functions. In addition, MathWorks and other manufacturers provide specialized toolkits that extend the functionality of MATLAB. There are also some toolkits available for free download online.

1.5 Punctuation in MATLAB

The table below lists some of the most used punctuation symbols in MATLAB. Arithmetic Operators in A.5 In MATLAB, the arithmetic operators (*, /, +, -, and) adhere to the linear algebraic convention. Two matrices, A and B, must be dimensionally right if we are to multiply them. In other words, A's number of columns and B's number of rows must match. We only need to multiply by A*B. It is crucial.

Punctuation	Usage
%	A percent sign denotes a comment line. Information after the % is ignored
…	Three periods denotes the continuation of a statement. Comment statements and variable names cannot be continued with this punctuation..
;	A semi-colon suppresses printing the contents of the variable to the screen. It also concatenates array elements along a column.
!	! An exclamation tells MATLAB to execute the following as an operating system command.
:	: The colon specifies a range of numbers. For example, 1:10 means the numbers 1 through 10. A colon in an array dimension accesses all elements in that dimension.
.	The period before an operator tells MATLAB to perform the corresponding operation on each element in the array

1.6 Building An Array

The data would be imported into MATLAB using load or another method discussed previously since the statistician or engineer will typically be using external data in an analysis. Simple arrays may occasionally need to be typed in order to test code, enter parameters, etc. Here, we discuss a few strategies for creating compact arrays. Keep in mind that you may concatenate arrays using technique as well.

Columns of elements (which can be arrays) are joined by commas or spaces. As a result, the following gives us a row vector. temp = [1, 4, 5]; or we can concatenate two column vectors a and b into one matrix, as follows temp = [a b];
The semi-colon tells MATLAB to concatenate elements as rows. So, we would get a column vector from this command: temp = [1; 4; 5];

1.7 Functions for Calculating Descriptive Statistics Function Summary

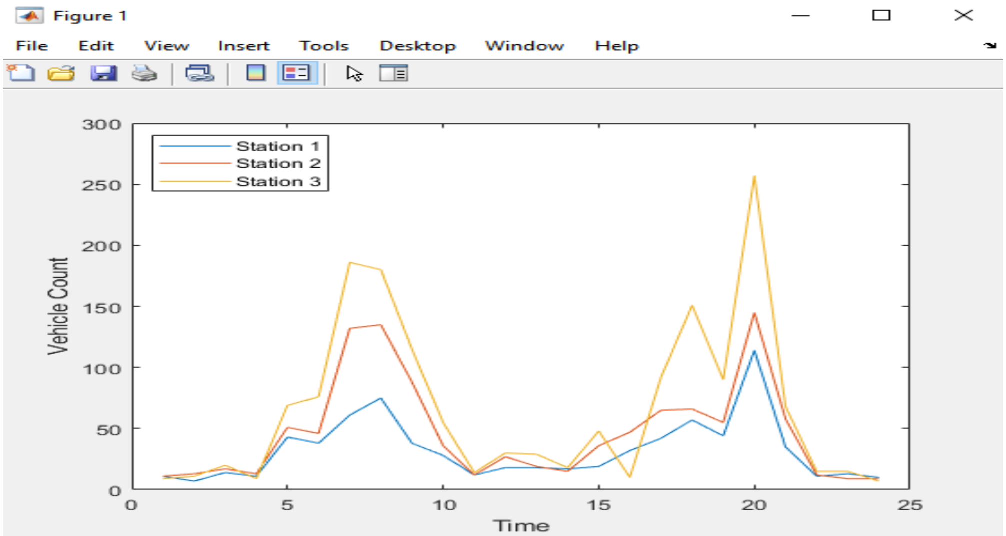
Function	Description
max	Maximum value
mean	Average or mean value
median	Median value
min	Smallest value
mode	Most frequent value

Function	Description
std	Standard deviation
var	Variance, which measures the spread or dispersion of the values

1.8 Calculating and Plotting Descriptive Statistics

1. Load and plot the data:

- load count.dat
- [n,p] = size(count);
- % Define the x-values
- t = 1:n;
- % Plot the data and annotate the graph
- plot(t,count)
- legend('Station 1','Station 2','Station 3','Location','northwest')
- xlabel('Time')
- ylabel('Vehicle Count')



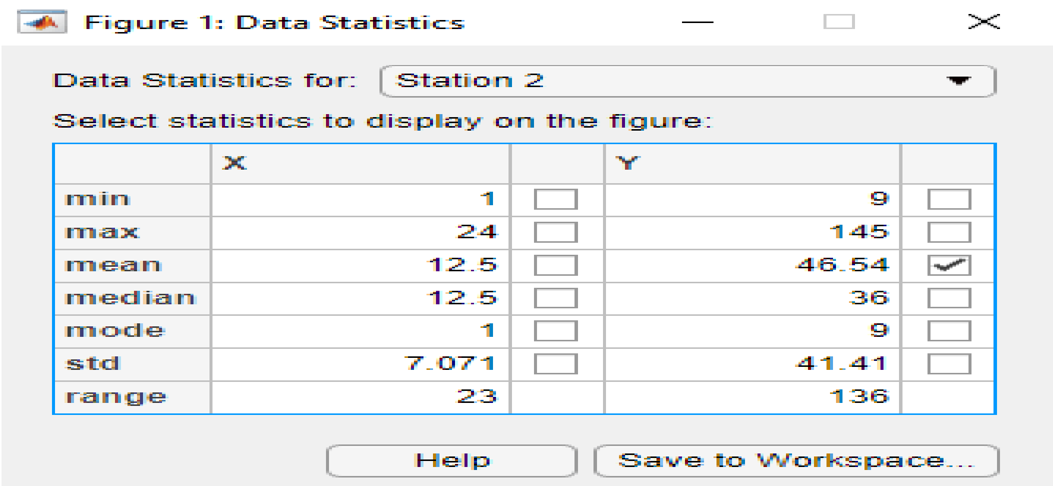
2. In the Figure window, select **Tools > Data Statistics**. The Data Statistics dialog box opens and displays descriptive statistics for the X- and Y-data of the Station 1 data set.
3. Select a different data set in the **Data Statistics for** list: Station 2.

This displays the statistics for the X and Y data of the Station 2 data set.

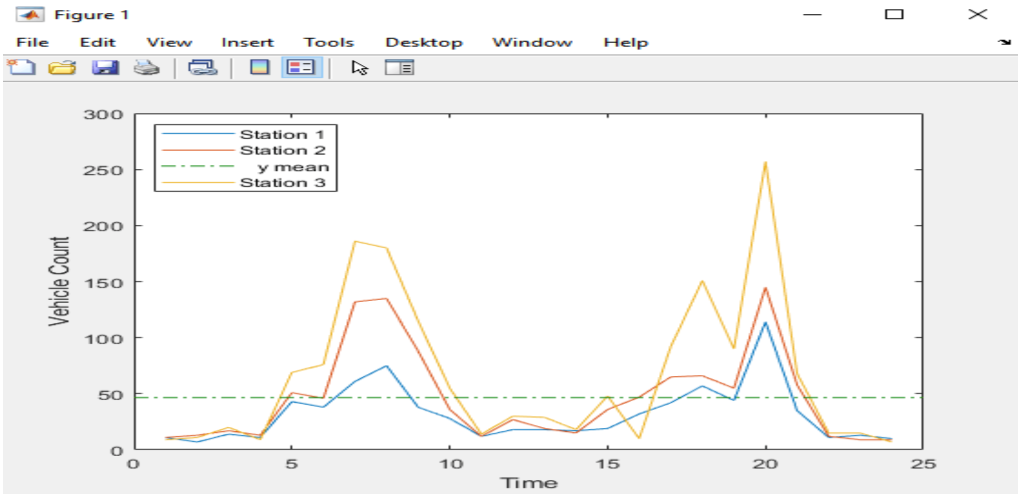
4. Select the check box for each statistic you want to display on the plot, and then click **Save to Workspace**.

For example, to plot the mean of Station 2, select the **mean** check box in the **Y** column

.



This plots a horizontal line to represent the mean of Station 2 and updates the legend to include this statistic.



REFERENCES

1. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. Journal of Geophysical Research 111, D12106. doi:10.1029/2005JD006548. Buckley, M.J., 1994.
2. Fast computation of a discretized thin-plate smoothing spline for image data. Biometrika 81, 247–258. Chatterjee, S., Hadi, A.S., 1986.
3. Influential observations, high leverage points, and outliers in linear regression. Statistical Science 1, 379–393. Craven, P., Wahba, G., 1978.
4. Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized crossvalidation. Numerische Mathematik 31, 377–403. Eilers, P.H., 2003. A perfect smoother. Anal. Chem. 75, 3631–3636. Garcia, D.,
5. Matlab functions in BioméCardio. http://www.biomecardio.com/matlab, 2009. Ref Type: Electronic Citation. Golub, G., Heath, M., Wahba, G., 1979.
6. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215–223. Hastie, T., Loader, C., 1993. Local regression: Automatic kernel carpentry.
7. Statistical Science 8, 120–129. Heiberger, R.M., Becker, R.A., 1992.
8. Design of an S function for robust regression using iteratively reweighted least squares. Journal of Computational and Graphical Statistics 1, 181–196. Hoaglin, D.C., Welsch, R.E., 1978.
9. The hat matrix in regression and ANOVA. The American Statistician 32, 17–22. Keller, H.B., 1965.
10. On the solution of singular and semidefinite linear systems by iteration. Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis 2, 281–290. Kennedy, J., Met Office Hadley Centre observations datasets. http://hadobs.metoffice.com/crutem3, 2007. Ref Type: Electronic Citation. Rousseeuw, P.J., Leroy, A.M., 1987.