MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.017/1.010 Computing and Data Analysis for Environmental Applications/Uncertainty in Engineering

---

Problem Set 1: Displaying data, MATLAB programming (Solutions provided at end of each problem)
Due: Thursday, Sept. 18, 2003

Please turn in a hard copy of your MATLAB source file as well as all printed outputs (tables, plots, etc.) required to solve the problem.

Problem 1

Go to the US census bureau web site at http://www.census.gov/statab/www/ranks.html to download data giving personal income and other information by state. To do this click on the "Personal Income Per Person" Excel link on the web site and save an Excel file containing the desired (1999) income data as well as various other things in your own working directory. Open the file in Excel and copy all lines in column 2: income per capita for each of the 50 states plus DC. Go into MATLAB and open a new file. Paste the income data in the file and save it as a .txt file. The file should consist of one column of numbers, with no text. Use the **find and replace …** option under **edit** to remove all commas (the commas confuse MATLAB). You can use another editor if you prefer. After you have simplified the file in this way save it.

You should write a MATLAB script in a file called **census.m** to process the census data. To bring the data into MATLAB you can use the load command:

```
load filename.txt
```

Now repeat the entire download process for the "Energy Consumption Per Person" link in the census web site, creating a new array called **energy** in your script. Pick two other variables of interest to you from the list provided on this site. In your script create arrays for each of these indicators with appropriate names, just as you did for **income** and **energy**. Be sure to divide totals by population if you want to compare per capita values.

Now include in your MATLAB script the ability to plot histograms and CDFs for each of the four data sets you downloaded from the census site. Also compute the mean, median, and variance for each data set. Plot a scatter diagram of energy consumption vs. income. Plot two other scatter diagrams of interest to you, selecting from the variables offered on the web site. Briefly indicate what each of these three scatter diagrams suggest to you.

In your discussion pay particular consideration to outliers that may violate a general trend.

Be sure to include labels on the plots you produce. The preferred approach is to create these labels from within the script, by using the **xlabel**, **ylabel**, and **title** functions (see MATLAB Help and the example from Recitation 1). Be sure also to turn in the MATLAB output giving the values you calculated for mean, median, and variance in addition to the plots and your program.

Problem 1 Solution

```
% 1.017/1.010 Problem Set 1 -- Problem 1
close all
clear all
load income.txt  % per capita income in 1996 dollars
load energy.txt  % per capita energy use in mil Btu
load fed_aid.txt % per capita federal aid in 1996 dollars
load unemployment.txt   % percent unemployment
mean(income)
median(income)
var(income)
mean(energy)
median(energy)
var(energy)
mean(fed_aid)
median(fed_aid)
var(fed_aid)
mean(unemployment)
median(unemployment)
var(unemployment)
% plot histograms and cdf's for each
figure
hist(income,6)
xlabel('Average Per Capita Income in 1996 USD')
ylabel('Number of States in Range')
title('Histogram of Per Capita Income for 50 States',…
' and DC')
figure
cdfplot(income)
xlabel('Average Per Capita Income in 1996 USD')
ylabel('F(income)')
title('Cumulative Distribution Function of Per Capita',…
'Income for the 50 States and DC')
figure
hist(energy,6)
xlabel('Per Capita Energy Use in mil Btu')
ylabel('Number of States in Range')
```

```
title('Histogram of Per Capita Energy Use for 50 States',…
' and DC')

figure
cdfplot(energy)
xlabel('Per Capita Energy Use in mil Btu')
ylabel('F(energy)')
title('Cumulative Distribution Function of Per Capita',…
' Energy Use for the 50 States and DC')
figure
hist(fed_aid,5)
xlabel('Per Capita Federal Aid in 1996 USD')
ylabel('Number of States in Range')
title('Histogram of Per Capita Federal Aid for 50 States',…
' and DC')
figure
cdfplot(fed_aid)
xlabel('Per Capita Federal Aid in 1996 USD')
ylabel('F(Aid)')
title('Cumulative Distribution Function of Per Capita',…
' Federal Aid for the 50 States and DC')
figure
hist(unemployment,6)
xlabel('Unemployment Rate [%]')
title('Histogram of Unemployment Rate for the 50 States',…
' and DC')
figure
cdfplot(unemployment)
xlabel('Unemployment Rate [%]')
ylabel('F(Unemployment)')
title('Cumulative Distribution Function of Unemployment
Rate for the 50 States and DC')
figure
plot(income,energy,'*')
title('Per Capita Energy Use vs Income')
xlabel('Income [1996 USD]')
ylabel('Energy [mil BTU]')
figure
plot(income,fed_aid,'o')
title('Per Capita Federal Aid vs Income')
xlabel('Income [1996 USD]')
ylabel('Federal Aid [1996 USD]')
figure
plot(income,unemployment,'x')
title('Unemployment Rate vs Per Capita Income')
xlabel('Income [1996 USD]')
ylabel('Unemployment Rate [%]')
```

Problem 2

Write a MATLAB script that computes tidal elevation h (in m) from the following three term series:

$$h(t) = 0.2Sin(0.5\omega t) + 3.0Sin(\omega t) + 1.0Sin(2\omega t)$$

where $\omega = 2\pi/24$ hr^{-1}. Define a vector **time** with 1001 values [0. 0.1 0.2 ….. 100.0.] spaced every 0.1 hours from $t = 0$ to $t = 100$ hrs. Use the vector capabilities of the MATLAB **sin** function to compute a vector of 1001 corresponding tidal elevations with a **single equation**. Read the MATLAB help discussion of the **sin(X)** function to make sure that you understand how it provides element-wise evaluations of vector arguments. Plot the tidal elevations from vs. time. Label both axes and include a plot title.

Problem 2 Solution:

```
% 1.017/1.010 Problem Set 1 -- Problem 2
close all
clear all
% compute tidal elevation
omega=2*pi/24;       % [hr^-1]
time=0:.1:100;
h=0.2*sin(0.5*omega*time)+3*sin(omega*time)+ …
sin(2*omega*time);
figure
plot(time,h)
title('Tidal Elevation vs Time')
xlabel('Time [hr]')
ylabel('Tidal Elevation [m]')
```