Table of Contents

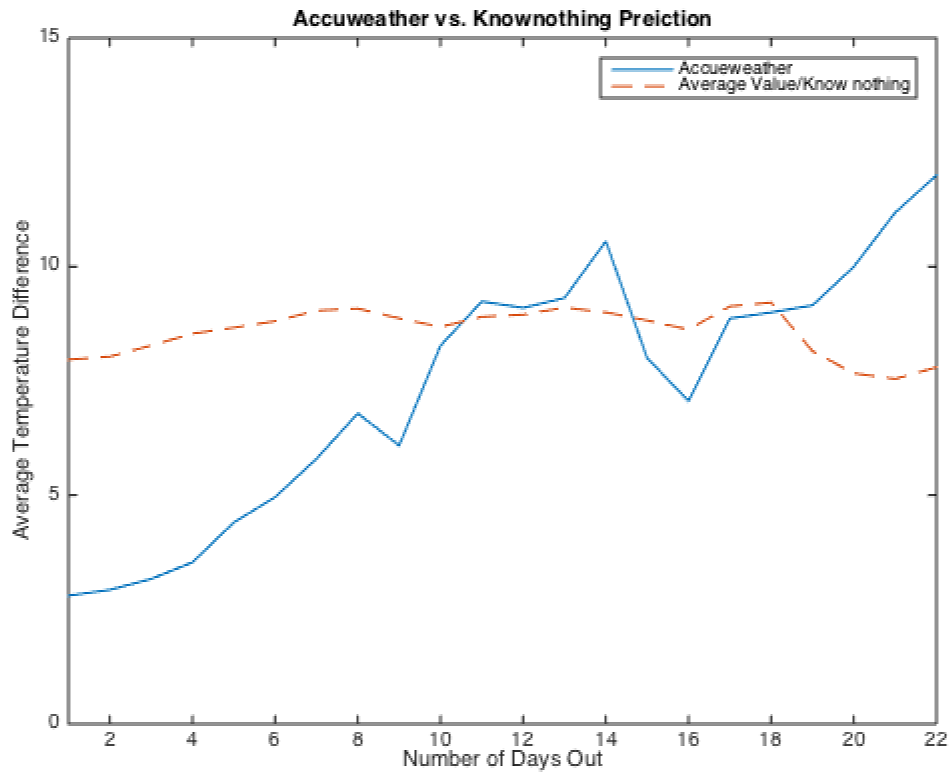

```
% Project 3
% Patrick Stepien
% 02/25/16
```

Problem 1

```
% In order to load the data into Matlab I first saved the excel file onto
% my desktop. I then used the Matlab tool "Import Data" found under the
% home tab to import the data into Matlab before renaming each vector
% to a more discriptive name.

load('Project_3_data');

% In order to then produce the data as a function of how many days out
% the prediction was made I turned it into a data structure. To do this
% I made the data structure 'janurary' and made 9 different fields for
% each type of data. Then I set each of the rows of each field equal to
% 31 day intervals. This creates it so that each row of each field
% is a 31 x 1 vector of all of the predictions for so many days out. For
% example the first row consists of all of the same day predictions, the
% second row the next day predictions, so on and so on.

k = 1;

for i = 1:31

janurary(i).AvgHi = AvgHi(k:k+30,1);
janurary(i).AvgLo = AvgLo(k:k+30,1);
janurary(i).High = High(k:k+30,1);
janurary(i).Low = Low(k:k+30,1);
janurary(i).Precip = Precip(k:k+30,1);
janurary(i).Snow = Snow(k:k+30,1);
janurary(i).TruHi = TruHi(k:k+30,1);
janurary(i).TruLo = TruLo(k:k+30,1);
janurary(i).TruPrecip = TruPrecip(k:k+30,1);

k = k+31;

end
```

Problem 2

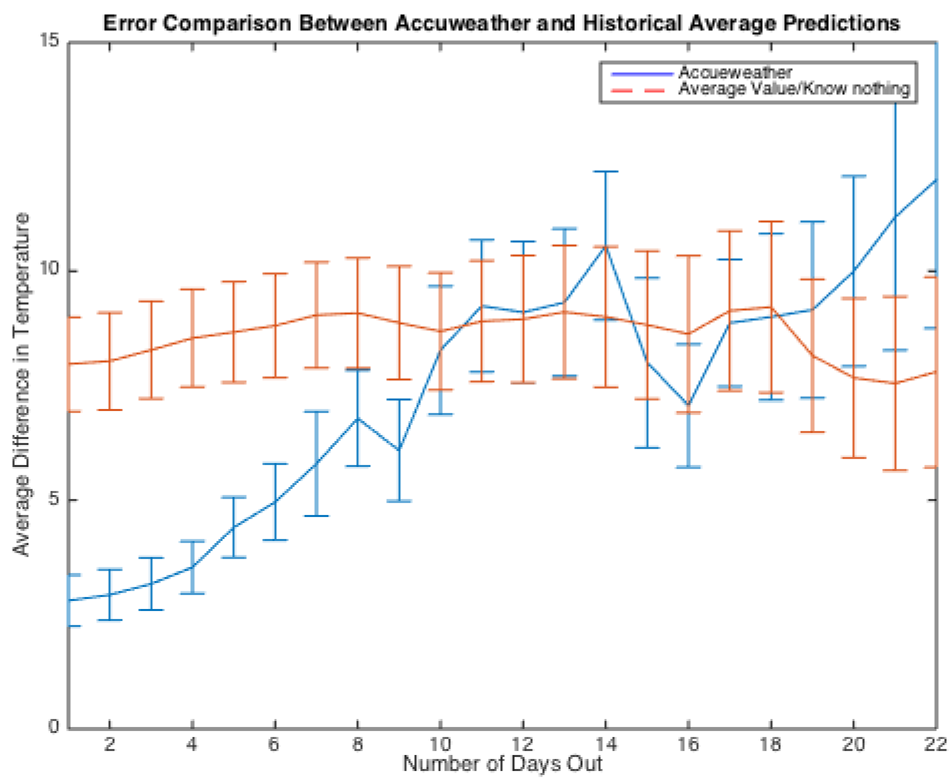
```
%The following for loop creates a second data structure named DiffTemp
%with two fields 'prediction' and 'average'. The prediction field holds
%the difference in the Accuweather predicted hightemps and the actual
%high temperature, while the average field holds the difference in the
%historical average high temperatures and the actual high temperatures.
%This is more of a random or know nothing prediction which can then be
%compared to the Accuweather predictions.

for i = 1:31
    h = i;
    j= 1;
  for k = h:31
    DiffTemp(i).prediction(j,1) = abs(janurary(j).High(k,1)-...
        janurary(j).TruHi(k,1));

    DiffTemp(i).average(j,1) = abs(janurary(j).AvgHi(k,1)-...
        janurary(j).TruHi(k,1));

    j = j+1;
  end
end

 AvgDiffTempprediction = zeros(31,1);
 AvgDiffTempaverage= zeros(31,1);

%Find the average of both the technology based and the know nothing
%predictions.
for i = 1:31
    AvgDiffTempprediction(i,1) = mean(DiffTemp(i).prediction);
    AvgDiffTempaverage(i,1) = mean(DiffTemp(i).average);
end

%Plot up the two predictions to graphically compare them.
figure(1)
plot([1:31],AvgDiffTempprediction,[1:31],AvgDiffTempaverage,'--')
axis([1,22,0,15])
xlabel('Number of Days Out');
ylabel('Average Temperature Difference')
title('Accuweather vs. Knownothing Preiction')
legend('Accueweather','Average Value/Know nothing')
```



Problem 3

```
%Find the Standard Deviation of the average of of both predictions and plot
%this on the graph in the form of error bars to see the trend of the error
%as the predicitons get further out.
stdDiffTempprediction = zeros(31,1);
stdDiffTempaverage = zeros(31,1);
for i = 1:31
   sigmaDiffTempprediction(i,1) = std(DiffTemp(i).prediction/sqrt((32-i)));
   sigmaDiffTempaverage(i,1) = std(DiffTemp(i).average/sqrt((32-i)));
end


figure(2)

plot([1:31],AvgDiffTempprediction,'b',[1:31],AvgDiffTempaverage,'r--')
axis([1,22,0,15])
xlabel('Number of Days Out')
ylabel('Average Difference in Temperature')
title(['Error Comparison Between Accuweather and Historical Average',...
       ' Predictions'])
legend('Accueweather','Average Value/Know nothing')

hold on
errorbar([1:31],AvgDiffTempprediction,sigmaDiffTempprediction,...
```

```
    sigmaDiffTempprediction)
errorbar([1:31],AvgDiffTempaverage,sigmaDiffTempaverage,...
    sigmaDiffTempaverage)
hold off

%Form observing this graph it becomes clear that as the predictions get
%further out the standard deviation between the predicted high temp. and
%the actual high temp gets larger and larger. This shows that as the
%predictions get further out they get significantly less reliable.
%This graph also shows that the Accuweather error becomes equal to the
%Historical Average error after about 7 days showing that the technology
%that Accuweather uses is really only helpful to about 7 days out.
```

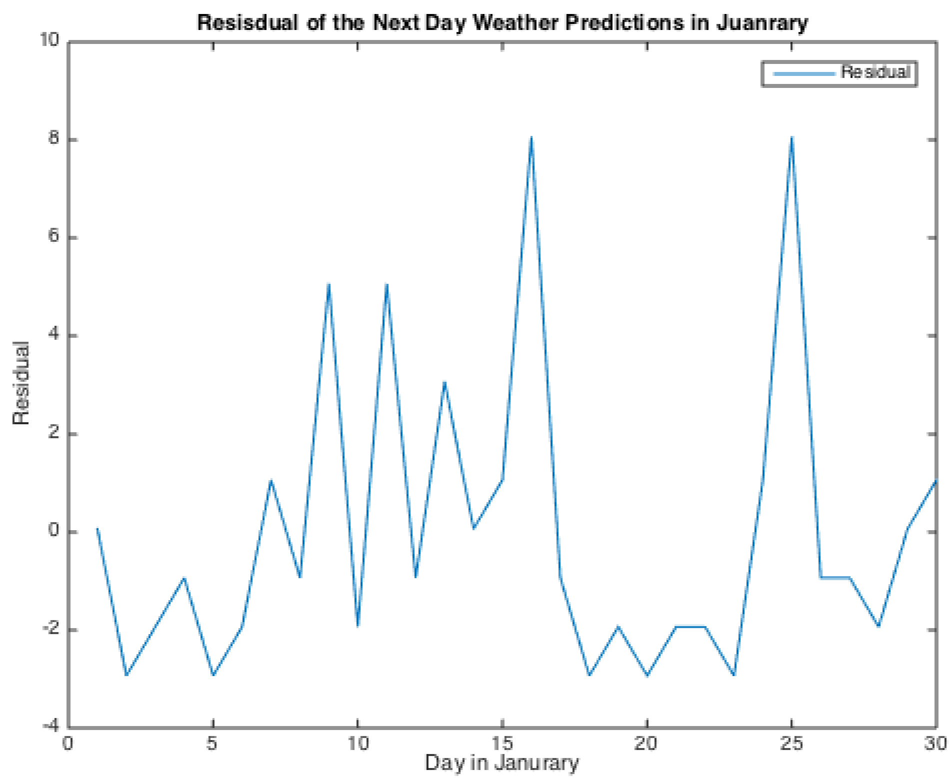


Problem 4

```
%Create a vector of just the 30 next day predicitions.
NextDayPred = DiffTemp(2).prediction;

%Calculate the Residuals each of these Predictions
Residual = NextDayPred-mean(NextDayPred);

figure(3)
plot(Residual)
xlabel('Day in Janurary')
ylabel('Residual')
title('Resisdual of the Next Day Weather Predictions in Juanrary')
```

```
legend('Residual')
%Based on this graph it would appear that the residuals are very random.
%This shows that the data is probably independent.

%Calculate the Variance and Covariance
Covar = dot(Residual(1:29,1),Residual(2:30,1))/(29-1);
Var = dot(Residual,Residual)/(30-1);

Ratio = Covar/Var;

disp(Ratio)

% As you can see the ratio between the Covariance and Variance is very
% small. This shows that the Covariance is very low compared to the
% Variance shows the days are not dependent on eachother and shows that
% the data is infact random.
```

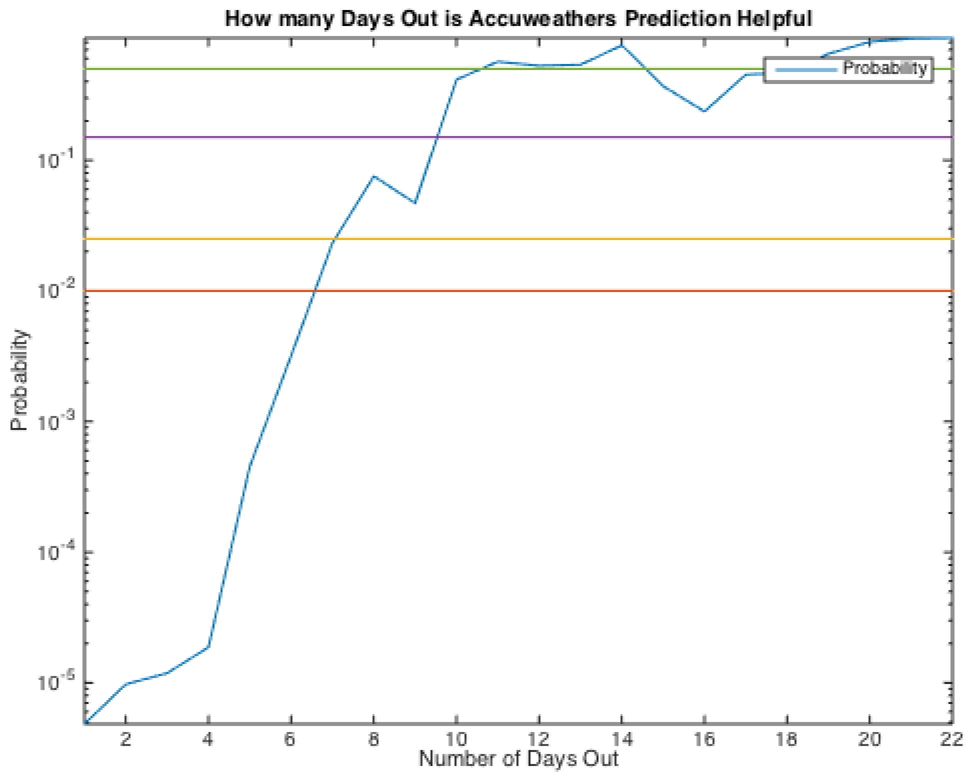
```
    0.0453
```



Problem 5

```
%In order to turn the results from part 3 into a probability I first
%calculated the T statistic and then used a normal cumulative distribution
%function.By plotting this with a semilogy scale I was properly able to
%plot the probablity that you are better off with the historical
```

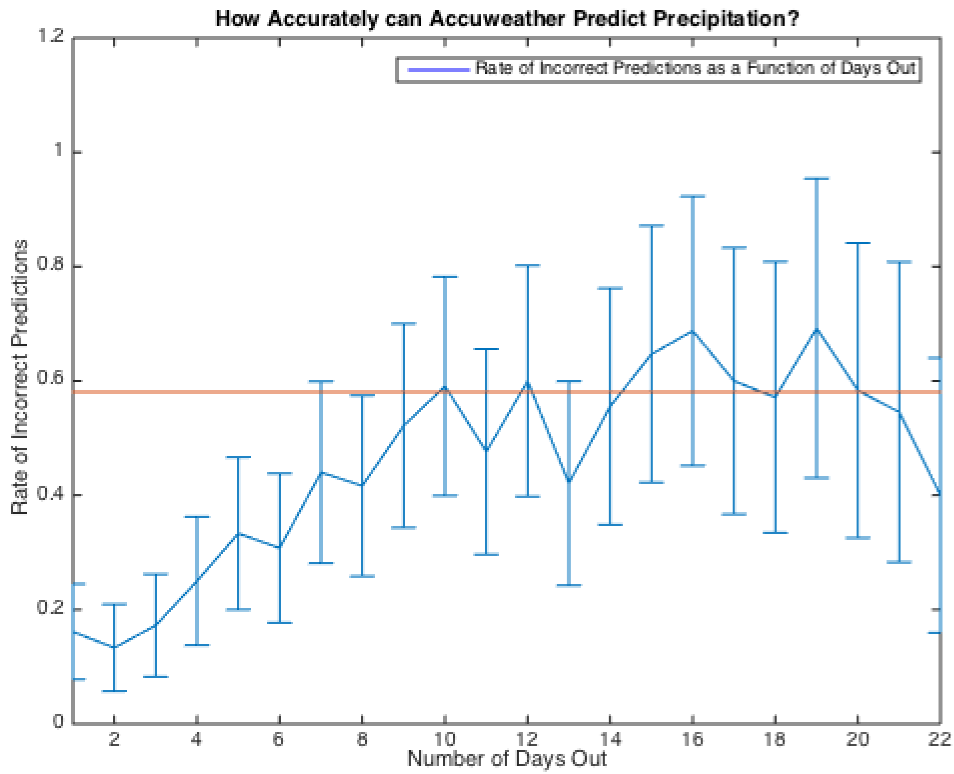
```
%average prediction rather than the Accuweather prediction.

sigmaTstat = (sigmaDiffTempprediction.^2 + sigmaDiffTempaverage.^2).^.5;

Tstat=(AvgDiffTempprediction-AvgDiffTempaverage)./sigmaTstat;

Tstat(31,1) = 0; %The error for the prediction will be 0 because there is
                 %Only one measurement for this value.

Prob = normcdf(Tstat,0,1);

% Create Lines at 1 percent, 2.5 percent, 15 percent and 50 percent.
days=linspace(1,22,22);
prob1 = 0.01*ones(1,22);
prob2 = 0.025*ones(1,22);
prob3 = 0.15*ones(1,22);
prob4 = 0.5*ones(1,22);

figure(4)

semilogy(Prob)
axis([1,22,0,inf]);

hold on
xlabel('Number of Days Out')
ylabel('Probability')
title('How many Days Out is Accuweathers Prediction Helpful')
legend('Probability')
plot(days,prob1,days,prob2,days,prob3,days,prob4)
hold off
% By observing this graph it becomes clear that the Accuweather prediction
% is in the 99% confindence up to about 6 days out. Once the prediction
% gets to about 10 days out the probability of it
% being more accurate than the Know-nothing prediction becomes a coinflip
% showing that the predictioin is more or less becomes useless.
```



Problem 6

```
%This for loop is neccessary because there is at least one value for
%precipitation that is over 1. What the for loop does is it creates a new
%datastructure Precip1 that will put a 1 for any day that the prediction
%on if it would precipitate or not was incorrect and a 0 for any day that
%the prediction was correct. This lets us see how accurate Accuweather
%predicted the percipitation as a function of how far in advanced they
%were predicting it.

for i = 1:31
    h = i;
    j= 1;
  for k = h:31
        if janurary(j).Precip(k,1)>1
           janurary(j).Precip(k,1)=1;
         Precip1(i).error(j,1) = abs(ceil(janurary(j).Precip(k,1))-...
          ceil(janurary(j).TruPrecip(k,1)));
        elseif janurary(j).TruPrecip(k,1)>1
               janurary(j).TruPrecip(k,1)=1;
            Precip1(i).error(j,1) = abs(ceil(janurary(j).Precip(k,1))-...
          ceil(janurary(j).TruPrecip(k,1)));
        else
            Precip1(i).error(j,1) = abs(ceil(janurary(j).Precip(k,1))-...
          ceil(janurary(j).TruPrecip(k,1)));
```

```
        end
        j = j+1;
    end
end

%Create a Rate matrix that is the sum of all the missed predictions
%divided by the total number of predictions for each prediction range.
N = zeros(31,1);
Rate = zeros(31,1);

for i = 1:31
    N(i,1) = sum(Precip1(i).error);
    Rate(i,1) = N(i,1)./(32-i);
end

for i = 1:31
    sigmaRate(i) = (Rate(i).^(1-Rate(i))/(32-i)).^.5;
end

averagefrac = (18/31)*ones(1,31); %There were 18 days in Janurary with
                                  %precipitation.

%By plotting the Rate of incorrect predictions up with errorbars and
%a line showing the average fraction of days in Janurary that it actually
%rained you can very well see that once again as the prediction gets
%further and further out it becomes less and less accurrate. At right
%around 10 days the graph crosses the average percentage of times it
%actually rained in Janurary showing that at 10 days out or more
%Accuweathers prediction of the precipitation is not a very good one.
figure(5)

plot([1:31],Rate,'b')
axis([1,22,0,1.2])
xlabel('Number of Days Out')
ylabel('Rate of Incorrect Predictions')
title('How Accurately can Accuweather Predict Precipitation?')
legend('Rate of Incorrect Predictions as a Function of Days Out')
hold on
errorbar([1:31],Rate,sigmaRate,sigmaRate)
plot([1:31],averagefrac)
hold off
```



Published with MATLAB® R2014b