Data analysis 6
Week 9 practical

Is most of the planet covered with clouds, on average, or not?
Is the liquid cloud coverage greater (or smaller), on average, than the ice cloud coverage?
Do months within an El-Nino year have a higher cloud coverage, on average, than those that are not in an El-Nino?
These are the questions to be addressed in this practical.

First download the `MODIS_satellite_data.zip` file from *Spreadsheets and data* and unzip it on your hard drive. MATLAB users will also want to download the function m-files in this weeks folder (`additional_matlab_files.zip`) and put them in the same folder as the satellite data.

6.1 Overview

In this practical you will test hypotheses about a sample mean and also about two sample means.

- The data used are real data of *cloud fraction* sampled using the MODIS satellite sensor.
- The MODIS sensor is on a satellite platform that is *polar orbiting* (see `http://www.youtube.com/watch?v=MLrPyhT_YNg`) and so it samples a small fraction of the globe at any one time.
- Nevertheless, eventually the sensor will sample data from ever point on the globe, many times during a month.
- *Cloud fraction* is the fraction of times a point on the globe is sampled to be cloudy to the times the same point is sampled.
- In this practical you are given 12 one-monthly datasets of global cloud fraction and asked to infer questions about the population mean cloud fraction from within and out of El-Nino years.
- You therefore need to calculate a global average for each month and use the 12 global averages as your sample data to test hypotheses against.

At the end, save your spreadsheet or MATLAB workspace (e.g. type: `save <filename>`).

6.2 Questions asked and statistics

1. Is the global total cloud fraction (liquid plus ice) from months within El-Nino years significantly greater than 0.5 (in other words is most of the planet covered with clouds)?

Here we are testing a hypothesis about a mean (as covered in the last practical). The statistic to be tested is therefore:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

with $N - 1$ ***degrees of freedom.***

2. Is the global liquid cloud fraction from months within an El-Nino year greater (or smaller) than the ice cloud fraction?
   Here we are testing a hypothesis about two means (as covered in the Monday lecture). The statistic to be tested is therefore:

$$\begin{aligned} S_p^2 &= \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} \\ t &= \frac{\bar{x_1} - \bar{x_2}}{\sqrt{S_p^2/n_1 + S_p^2/n_2}} \end{aligned}$$

   with $n_1 + n_2 - 2$ ***degrees of freedom.***
3. Is the global total cloud fraction (liquid plus ice) from months within El-Nino years significantly greater than those not in El-Nino years?
   Here we are testing a hypothesis about two means (as covered in the Monday lecture). The statistic to be tested is therefore as in question 2.

For further information see the sections on either Excel or MATLAB below.

6.3 Excel

If you click on the files they will open up in Excel. You'll notice the data start in cell `A1` and end in row 181 and column MW. The first row is the value of *longitude* that the data are taken at and the first column is the value of *latitude*. The rest of the data are *cloud fraction* values at those *longitude* and *latitude*.

1. For the first question:
   - You can take the global mean of a particular dataset by typing: `=average(B2:MW181)` in an empty cell.
   - You can also reference a different work book to where you calculate the average by opening up another workbook and calculating the means by using the mouse to select the numbers you want.
   - Calculate the means for all csv files that begin with the letter `A` and end in `_all.csv`; there should be 12 means in total. These are the total cloud fractions for the El-Nino year.
   - You should have all you need to test the hypothesis now.
2. For the second question:

- Repeat the above process for the files starting with `A` and ending with `_liquid.csv` and `_ice.csv`; there should be two lots of 12 means (24 values in total).
- You should have everything you need to test the hypothesis that the liquid cloud fraction is different to the ice cloud fraction.

3. For the third question:
    - Repeat the above process for the files starting with `C` (this is the non El-Nino year) and ending with `_all.csv`; there should be 12 additional means to those you generated in question 1 (24 values in total).
    - You should have everything you need to test the hypothesis that the cloud fraction is different in El-Nino years is different to the cloud fraction in non El-Nino years.

6.4 MATLAB

Question 1: To test whether more of the earth is covered by clouds than is not, take a look at the following code.

```
function [means]=process_cloud_fraction2
% Read in the cloud fraction data and calculate a global mean for each
% month. Is the global mean >50%?

d=dir('A*_all.csv');

N=length(d); % sample size
means=zeros(size(d));
for i=1:length(d)
    dat=csvread(d(i).name);
    cloud_fraction=dat(2:end,2:end);
    means(i)=nanmean(cloud_fraction(:));

%     % problem:
%     [longs,lats]=meshgrid(dat(1,2:end),dat(2:end,1));
%     weighted=cloud_fraction.*cos(lats.*pi./180);
%     means(i)=nanmean(weighted(:));
end

% Is mean cloud cover >0.5? test at the 5% significance level
t_statistic=abs((mean(means)-0.5)./(std(means)./sqrt(N)))

t_critical=abs(tinv(0.05,N-1))

if(t_critical>t_statistic)
    disp(['Accept the null hypothesis: ', ...
        'there is no difference between 0.5 and the sample data']);
else
    disp(['Reject the null hypothesis: ', ...
        'there is a difference between 0.5 and the sample data']);
end
```

This code is downloadable from Blackboard in the folder for this weeks practical. If you put it on your hard disk in the same directory as the csv files you will be able to run it from the MATLAB commandline by typing:

```
[means]=process_cloud_fraction2;
```

and `means` will contain the 12 global means.

Question 2: To test whether more of the earth is covered by liquid clouds than ice clouds, take a look at the following code. *You may need to alter the significance level for your test.*

```
function [means1,means2]=process_cloud_fraction3
% Read in the cloud fraction data and calculate a global mean for each
% month. Is the global mean fraction of liquid clouds greater than ice
% clouds?

d1=dir('A*_liquid.csv');
d2=dir('A*_ice.csv');

N=length(d1); % sample size
means1=zeros(size(d1));
means2=zeros(size(d1));
for i=1:length(d1)
    dat1=csvread(d1(i).name);
    cloud_frac_liquid=dat1(2:end,2:end);
    means1(i)=nanmean(cloud_frac_liquid(:));

    dat2=csvread(d2(i).name);
    cloud_frac_ice=dat2(2:end,2:end);
    means2(i)=nanmean(cloud_frac_ice(:));

end

% Is mean liquid cloud cover > mean ice cloud cover?
% test at the 5% significance level
Sp2=(N-1).*std(means1).^2+(N-1).*std(means2).^2;
Sp2=Sp2./((N-1)+(N-1));
t_statistic=abs((mean(means1)-mean(means2))./(sqrt(Sp2./N+Sp2./N)) )

t_critical=abs(tinv(0.05,N+N-2))

if(t_critical>t_statistic)
    disp(['Accept the null hypothesis: ', ...
        'there is no between the fraction of liquid and ice clouds']);
else
    disp(['Reject the null hypothesis: ', ...
        'there is a difference between the fraction of ',...
        'liquid and ice clouds']);
end
```

This code is downloadable from Blackboard in the folder for this weeks practical. If you put it on your hard disk in the same directory as the csv files you will be able to run it from the MATLAB command line by typing:

```
[means1,means2]=process_cloud_fraction3;
```

`means1` and `means2` will contain the 12 liquid and ice fraction global means respectively.

<u>*Question 3:*</u> 2006-7 was an El-Nino year. To test whether the cloud fraction is different in months of an El-Nino year vs months not in an El-Nino year look at the following code. *You may need to alter the significance level for your test.*

```
function [means1,means2]=process_cloud_fraction4
% Read in the cloud fraction data and calculate a global mean for each
% month. Is the global mean fraction of liquid clouds greater than ice
% clouds?

d1=dir('A*_all.csv');
d2=dir('C*_all.csv');

N1=length(d1); % sample size
N2=length(d2); % sample size
means1=zeros(size(d1));
means2=zeros(size(d2));
for i=1:length(d1)
    dat1=csvread(d1(i).name);
    cloud_frac_2006=dat1(2:end,2:end);
    means1(i)=nanmean(cloud_frac_2006(:));
end
for i=1:length(d2)
    dat2=csvread(d2(i).name);
    cloud_frac_2011=dat2(2:end,2:end);
    means2(i)=nanmean(cloud_frac_2011(:));
end

% Is mean liquid cloud cover > mean ice cloud cover?
% test at the 10% significance level
Sp2=(N1-1).*std(means1).^2+(N2-1).*std(means2).^2;
Sp2=Sp2./((N1-1)+(N2-1));
t_statistic=abs((mean(means1)-mean(means2))./(sqrt(Sp2./N1+Sp2./N2)) )

t_critical=abs(tinv(0.10,N1+N2-2))

if(t_critical>t_statistic)
      disp(['Accept the null hypothesis: ', ...
            'there is no between the first year and the second year']);
else
      disp(['Reject the null hypothesis: ', ...
            'there one year is cloudier than the other']);
end
```

This code is downloadable from Blackboard in the folder for this weeks practical. If you put it on your hard disk in the same directory as the csv files you will be able to run it from the MATLAB command line by typing:

```
[means1,means2]=process_cloud_fraction4;
```

`means1` and `means2` will contain the 12 El-Nino means and non El-Nino global means respectively.

6.5 Fancy stuff in MATLAB

Display a pseudo colour plot of the cloud fraction from September 2012:

```
% Display the cloud fraction for September 2012
dat=csvread('C201209_all.csv');

% parse 'dat' into long, lat and cloud fraction.
long1=dat(1,2:end); % First row of data
lat1=dat(2:end,1);  % First column of data
cloud_fraction=dat(2:end,2:end); % rest of the data

figure('name','Cloud fraction')
pcolor(long1,lat1,cloud_fraction);shading flat;
colormap gray;
xlabel('longitude');
ylabel('latitude');
title('Cloud fraction for September 2012')
colorbar

% Display coast line data
load coast
hold on;
plot(long,lat,'color',[1 1 1]); % plot in white [1 1 1]
axis equal;axis tight;
% Print the figure as a png
print -dpng pcolor_cf.png
```

Display a contour plot of the cloud fraction from September 2012:

```
% Display the cloud fraction for September 2012
dat=csvread('C201209_all.csv');

% parse 'dat' into long, lat and cloud fraction.
long1=dat(1,2:end); % First row of data
lat1=dat(2:end,1);  % First column of data
cloud_fraction=dat(2:end,2:end); % rest of the data

figure('name','Cloud fraction')
[c,h]=contourf(long1,lat1,smoothn(cloud_fraction,[7 7]));
clabel(c,h);
colormap gray;
xlabel('longitude');
ylabel('latitude');
title('Cloud fraction for September 2012')
colorbar

% Display coast line data
load coast
hold on;
plot(long,lat,'color',[1 1 1]); % plot in white [1 1 1]
axis equal;axis tight;

% Print the figure as a png
print -dpng contour_cf.png
```