**Lab #2 Linear regression, spatial relationship and autocorrelation**

In this lab we will use rainfall data and DEM to calculate mean centers and weighted centers. In addition, you will learn how to use OLS and spatial correlation tool to analyze the spatial data distribution.

**What you need to submit:**

Hint: Copy and paste the questions below into a word document and submit it to blackboard

**Lab 2: Answer Sheet**  
Name:

**Part 1: Mean Center Tool**

**Question 1:** What is the difference between two mean centers? Briefly describe why the difference happens. **Attach the screen shot of your map with two centers here.** Please change the symbology for the two centers and add labels (optional) so that I could distinguish them.

**Part 2: Ordinary Least Squares Tool**

**Question 2:** Based on the attribution table, we have three new fields. Estimated, residual and stdResid. Attach your screenshot here.

Explanation about some key concepts:

**A residual** is the vertical distance between a data point and the regression line. Each data point has one residual. They are positive if they are above the regression line and negative if they are below the regression line. If the regression line actually passes through the point, the residual at that point is zero.

**A standardized residual** is the residual divided by its standard deviation. A general rule of thumb for figuring out what the standardized residual means, is:

If the residual is less than -2, the subject`s observed frequency is less than the expected frequency. (learn about observed / expected frequency <https://cnx.org/contents/05bErS_X@2/What-is-an-expected-frequency> )

Greater than 2 and the observed frequency is greater than the expected frequency.

**Question 3:** Use **Select By Attributes** to find out how many StdResid are greater than 2 and how many are lower than -2. Briefly describe the correlation between elevation and precipitation.

**Question 4:**

p-value:

z-score:

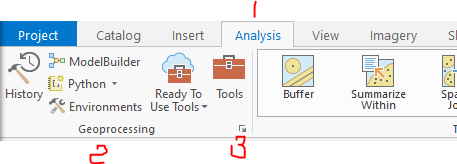
Moran's Index:

The rainfall distribution of this region is clustered or dispersed:

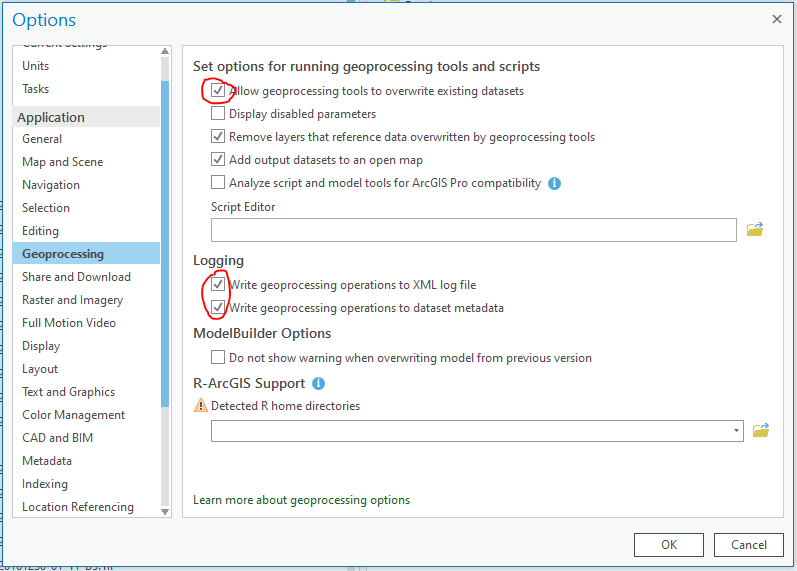
**Part 1. Mean Center Tool**

In this part, we will use spatial statistics tools to calculate mean center and precipitation weighted center for a set of features.

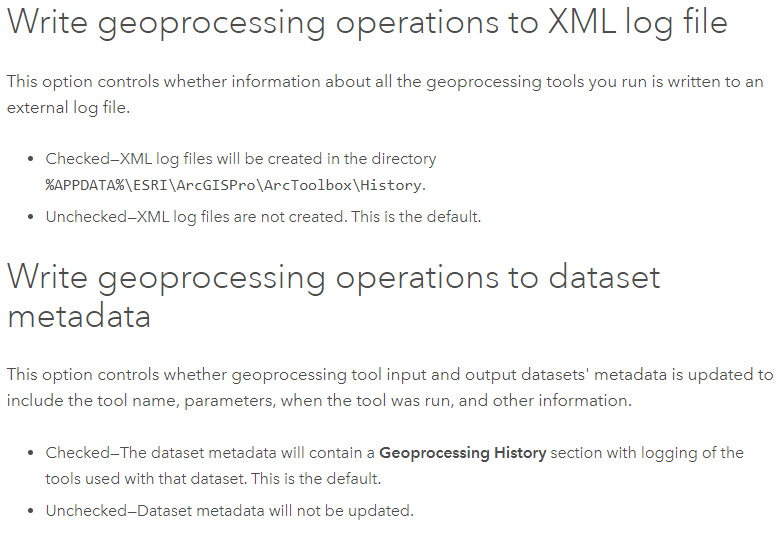
First, a few housekeeping items. Let’s take the training wheels off ArcGIS Pro and make sure we can overwrite data. Click Analysis > Geoprocessing > Geoprocessing Options.



* Make sure selected “Allow geoprocessing tools to overwrite existing datasets”.

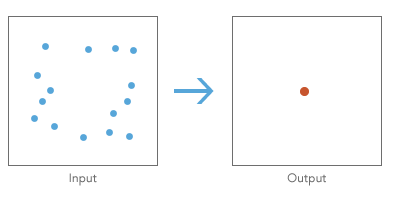


* To write geoprocessing operations to log files, you have two options. You could check either or both.
* For detailed information, please refer [Geoprocessing options—ArcGIS Pro | Documentation](https://pro.arcgis.com/en/pro-app/latest/help/analysis/geoprocessing/basics/geoprocessing-options.htm)



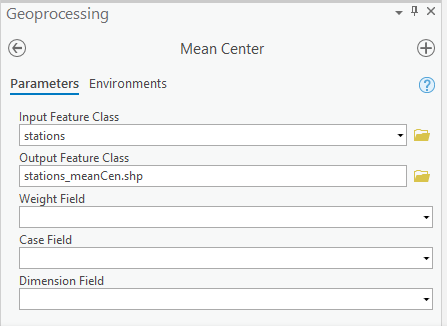
Let`s check the contents of lab2 folder in Catalog. There should be one raster dataset (dem.asc) file and one shapefile (stations.shp). Add these two files to a new map and change the data frame name as “lab02”.

We will identify the mean center for the stations. The tool in ArcGIS Pro is called Mean Center. The mean center is a point constructed from the average x, y and if available, z values for the input feature centroids. As the following illustration shows:



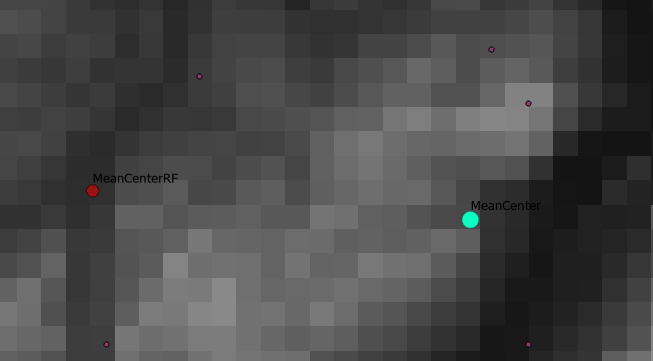
Go to Analysis > Geoprocessing > Tools, search Mean Center in Spatial Statistics Tools. In the following labs, we will use this search bar frequently.

Input Feature Class should be stations, Customize the Output Feature Class as stations\_meanCen.shp and store the output under lab02 folder. Leave following fields empty and click Run (on the bottom right corner).



On the Table of Contents (TOC), the stations\_meanCen layer should be added as the result.

If we give the Weight Field as Rainfall and a new name stations\_mean\_CenRain to the result layer, we will obtain the precipitation weighted center. Re-run the Mean Center tool and observe the difference between two mean centers. You can refer to the following image to check your own map. Answer the question #1.



**Part 2. Ordinary Least Squares Tool**

For the second part, we will extract elevation at the stations and then do a linear regression with precipitation and elevation using the OLS tool.

To extract elevation at the station points, we need not only the shapefile per se, but the dem.asc file. Go to Geoprocessing Tools and search **Extract Values to Points**. The input point features is stations, the input rasters is dem.asc. The Output point features can be st\_ele. Leave **Interpolate values at the point locations** unchecked. You can click the small icon besides and learn about how this affects the image. Click run.

Then open the attribute table of the created layer and you should view a new column named RASTERVALU.

Next, let do a linear regression with Ordinary Least Squares (OLS) tool. The goal is to estimate and predict precipitation based on elevation. Generally, OLS is to generate predictions or to model a dependent variable in terms of its relationships to a set of explanatory variables. Search the tool and open it. Stations should be given in Input Feature Class, Unique ID is ID field; Output Feature Class is stations\_lr.shp. Dependent Variable is Rainfall. Explanatory Variables is the elevation data you calculated from the above step (e.g. st\_ele). You may also give the name for output report, and save it under your lab02 folder. Click Run and stations\_lr layer should automatically appear in the TOC.

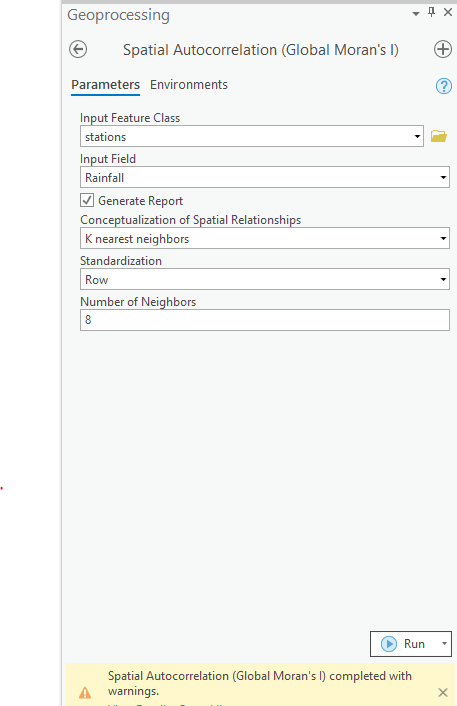
Note that the Dependent variable is the numeric field containing values for which you are trying to predict, while the explanatory variables is a list of fields (in our case just the elevation field) representing input variables in your regression model.

Open the attribute table of stations\_lr and answer the question #2 & 3.

**Part 3**

We will use Spatial Autocorrelation function to identify whether the rainfall of this region is dispersed or clustered.

Go to **Spatial Autocorrelation** (Global Moran`s 1), **Input Feature Class** is stations. **Input Field** is Rainfall. Check **Generate Report**. Select K nearest neighbors as Conceptualization of Spatial Relationships. Row as standardization. 8 as the number of Neighbors (3\*3 matrix).



Click Run and click View Details. Open the report and answer Question#4.