Assignment 2 (50 pts)

You may complete the assignment using your programming language of choice. Feel free to use built in functions but make sure you have read the documentation about these functions and are confident they are indeed conducting the calculations you intend. Please submit your assignment and the code used to generate any results by uploading the files to Canvas by the assignment due date. I should be able to re-create all of your results given the scripts you have provided me, this includes any data import commands. For your data loading, please ensure that your path is a relative path in the form '../../data/filename.ncl'. Your code should be well commented so that others can easily understand what has been done and marks may be removed from your assignment if this is not the case.

1. (20 pt) In this question we will explore the central limit theorem by testing different N sizes necessary for time averaged data to approach a Gaussian distribution. We will use the same SC_data_1980_2015.xlsx dataset as in Assignment 1. Only consider days with measureable precipitation (ie. set all values of 0 and -1 equal to missing values). Consider all values throughout the year.

a) For precipitation, compute time averages over a series of averaging periods: daily (no further averaging required), pentad (5-day), and monthly. After each set of averages, use a method of moments to fit a Gaussian distribution to the data. Plot histograms and fitted Gaussian distributions for each averaging period. Include the values of the fitted distribution parameters μ and σ in your title. Describe the impact of time averaging on the precipitation data. Note: The x-axis limits should be different between the plots.

b) For each data averaging period, standardize the data. Create Q-Q plots to qualitatively compare the data to a standard Gaussian distribution. Describe any differences you see between the fitted and empirical distributions.

c) Use a Chi-square goodness-of-fit test to quantitatively assess the goodness-of-fit to a standard Gaussian for each averaging period. What do these test results tell you about the averaging size N required to satisfy the central limit theorem?

2. (10 pt) For the following situations identify the significance test you would implement and why. This should include a description of the test statistic, null hypothesis (H0), and null distribution or method for obtaining the null distribution.

a) You believe that strong (Cat 5) hurricanes are more likely when the potentially intensity index of Emmanuel 1986 is high. The potential intensity data is available for all September Atlantic basin storms in the past 30 years. You compute the mean potential intensity for all 21 September Category 5 storms in the Atlantic basin and find that it is higher than the average potential intensity during a hurricane. You would like to know if these results are significant.

b) You went skiing 10 times at Vail this winter (you live in an alternate universe where Covid never happened). The average temperature on these 10 days was 35F, and the standard deviation of 10 daily temperatures is 5F. You know that the climatological mean winter temperature for Vail is 32F. You can assume that temperature is normally distributed. Your lift operator thinks this is a sign of climate change, is it?

3. (20 pt) In the following question you will be comparing precipitation statistics between two climate model simulations with radiative forcing from the early (2010-2019) or late (2090-2099) twentieth century. These data can be found in the class Box folder with the following names: FRCP85C5CN.001.2006-2019.CNTRL.cam.h1.PRECT.20060101-20191231.nc and FRCP85C5CN.001.2086-2099.CNTRL.cam.h1.PRECT.20860101-20991231.nc respectively.

a) Prior to beginning your analysis there are several data processing steps to complete. Firstly, we will be ignoring the first 4 years of data for spin up, and so you should only keep years 2010-2019 and 2090-2099. You will also only be analyzing the data over North America and so can limit your data spatially to 20-50°N and 130-65°W. The units of precipitation are in m/s and should be converted to mm/day. Note: It is good practice to read the metadata for your file prior to loading. Ex. Use the function ncdisp() in matlab.

b) In our analysis we will be examining two precipitation metrics, the median and interquartile range. Compute and plot maps of the median for the early simulation (2010-2019), late simulation (2090-2099), and their difference (2090-2099 minus 2010-2019). Repeat for the interquartile range. Note: These plots should be created using a suitable mapping projection (ex. Lambert), include latitude and longitude labels, coastlines, colorbars, and titles.

c) In order to determine whether the differences between the two simulations are statistically significant, you will employ a permutation test. Explain why this is an appropriate test for your data and test statistics.

d) Conduct the permutation tests at each grid point for both the median and IQR. For this test, use 1000 permutations and an α level of 95%. Add this test to your plot either by applying stippling to significant points or by masking out any points that are not significant. If you choose to mask out your points, be sure that it is possible to distinguish between small values and insignificant values. Note: the calculation may be time consuming.