**ATOC7500 – Application Lab #1**

**Significance Testing Using Bootstrapping and Z/T-tests**

**in class Monday August 31 and Wednesday September 2, 2020**

**Notebook #1 – Statistical significance using Bootstrapping**

**ATOC7500\_applicationlab1\_bootstrapping.ipynb**

**LEARNING GOALS:**

1) Use an ipython notebook to read in csv file, print variables, calculate basic statistics, do a bootstrap, make histogram plot

2) Hypothesis testing and statistical significance testing using bootstrapping

**DATA and UNDERLYING SCIENCE:**

In this notebook, you will analyze the relationship between Tropical Pacific Sea Surface Temperature (SST) anomalies and Colorado snowpack. Specifically, you will test the hypothesis that December Pacific SST anomalies driven by the El Nino Southern Oscillation affect the total wintertime snow accumulation at Loveland Pass, Colorado. When SSTs in the central Pacific are anomalously warm/cold, the position of the mid-latitude jet shifts and precipitation in the United States shifts. This notebook will guide you through an analysis to investigate the connections between December Nino3.4 SST anomalies (in units of °C) and the following April 1 Loveland Pass, Colorado Snow Water Equivalence (in units of inches). Note that SWE is a measure of the amount of water contained in the snowpack. To convert to snow depth, you multiply by ~5 (the exact value depends on the snow density).

The Loveland Pass SWE data are from:

<https://www.wcc.nrcs.usda.gov/snow/>

The Nino3.4 data are from:

<https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/>

**Questions to guide your analysis of Notebook #1:**

For full credit: write answers to the questions and then upload this document to your github along with notebook #1 (including any edits that you make).

1) Composite Loveland Pass, Colorado snowpack. Fill out the following table showing the April 1 SWE in all years, in El Nino years (conditioned on Nino3.4 being 1 degree C warmer than average), and in La Nina years (condition on Nino3.4 being 1 degree C cooler than average).

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mean SWE** | **Std. Dev. SWE** | **N (# years)** |
| **All years** | 16.33 | 4.22 | 81 |
| **El Nino Years** | 15.29 | 4.0 | 16 |
| **La Nina Years** | 17.78 | 4.11 | 15 |

2) Use hypothesis testing to assess if the differences in snowpack are statistically significant. Write the 5 steps. Test your hypothesis using bootstrapping.

1. State the significance level.

2. State the null hypothesis and the alternative.

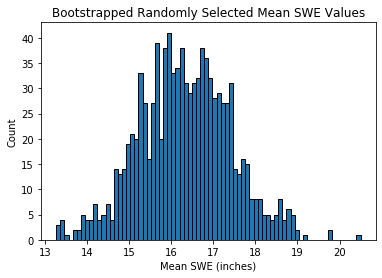
3. State the statistics to be used.

4. The null hypothesis is rejected if or if

5. Evaluate the statistics.

Instructions for bootstrap: Say there are N years with El Nino conditions. Instead of averaging the Loveland SWE in those N years, randomly grab N Loveland SWE values and take their average. Then do this again, and again, and again 1000 times. In the end you will end up with a distribution of SWE averages in the case of random sampling, i.e., the distribution you would expect if there was no physical relationship between Nino3.4 SST anomalies and Loveland Pass SWE.

1. Plot a histogram of this distribution and provide basic statistics describing this distribution (mean, standard deviation, minimum, and maximum).



Bootstrap mean: 16.32

Bootstrap standard deviation: 1.09

Bootstrap minimum value: 13.26

Bootstrap maximum value: 20.5

1. Quantify the likelihood of getting your value by chance alone using percentiles of this bootstrapped distribution. What is the probability that differences between the El Nino composite and all years occurred by chance? What is the probability that differences between the La Nina composite and all years occurred by chance?

The probability that the snowfall was lower during El Niño by chance is 15.95%.

The probability that differences between El Niño composites and all years occurred by chance is 34.77%.

The probability that the snowfall was higher during La Niña by chance is 9.01%.

The probability that differences between La Niña composites and all years occurred by chance is 18.03%.

3) Test the sensitivity of the results obtained in 2) by changing the number of bootstraps, the statistical significance level, or the definition of El Nino/La Nina (e.g., change the temperature threshold so that El Nino is defined using a 0.5 degree C temperature anomaly or a 3 degree C temperature anomaly). In other words – play and learn something about the robustness of your conclusions.

Reducing number of Bootstrap times from 1000 to 100 does not result in significant differences in the probabilities.

* The probability that the snowfall was lower during El Niño by chance is 15.57%.
* The probability that differences between El Niño composites and all years occurred by chance is 31.13%.
* The probability that the snowfall was higher during La Niña by chance is 7.47%.
* The probability that differences between La Niña composites and all years occurred by chance is 14.93%.

Reducing number of Bootstrap times from 1000 to 10 results in significant differences in the probabilities of La Niña event, but not for El Niño event.

* The probability that the snowfall was lower during El Niño by chance is 14.71%.
* The probability that differences between El Niño composites and all years occurred by chance is 29.41%.
* The probability that the snowfall was higher during La Niña by chance is 32.27%.
* The probability that differences between La Niña composites and all years occurred by chance is 64.53%.

Reducing the temperature threshold for El Niño and La Niña by 0.5 deg Celsius results in significant differences in the probabilities of La Niña event, but not for El Niño event.

* The probability that the snowfall was lower during El Niño by chance is 16.4%.
* The probability that differences between El Niño composites and all years occurred by chance is 32.79%.
* The probability that the snowfall was higher during La Niña by chance is 1.75%.
* The probability that differences between La Niña composites and all years occurred by chance is 3.5%.

Based on this “sensitivity” test, statistics obtained for El Niño are more robust than those obtained from La Niña.

4) Maybe you want to see if you get the same answer when you use a t-test… Maybe you want to set up the bootstrap in another way?? Another bootstrapping approach is provided by Vineel Yettella (ATOC Ph.D. 2018). Check these out and see what you find!!

Both the t-test and the modified bootstrap method reach the same conclusion when varying the temperature anomaly threshold that defines El Niño.

**Notebook #2 – Statistical significance using z/t-tests**

**ATOC7500\_applicationlab1\_ztest\_ttest.ipynb**

**LEARNING GOALS:**

1) Use an ipython notebook to read in a netcdf file, make line plots and histograms, and calculate statistics

2) Calculate statistical significance of the changes in a normalized mean using a z-statistic and a t-statistic

3) Calculate confidence intervals for model-projected global warming using z-statistic and t-statistic.

**DATA and UNDERLYING SCIENCE:**

You will be plotting *munged* climate model output from the Community Earth System Model (CESM) Large Ensemble Project. The Large Ensemble Project includes a 42-member ensemble of fully coupled climate model simulations for the period 1920-2100 (*note: only the original 30 are provided here*). Each individual ensemble member is subject to the same radiative forcing scenario (historical up to 2005 and high greenhouse gas emission scenario (RCP8.5) thereafter), but begins from a slightly different initial atmospheric state (created by randomly perturbing temperatures at the level of round-off error). In the notebook, you will compare the ensemble remembers with a 2600-year-long model simulation having constant pre-industrial (1850) radiative forcing conditions (perpetual 1850). By comparing the ensemble members to each other and to the 1850 control, you can assess the climate change in the presence of internal climate variability.

**More information on the CESM Large Ensemble Project can be found at:**

<http://www.cesm.ucar.edu/projects/community-projects/LENS/>

**Questions to guide your analysis of Notebook #2:**

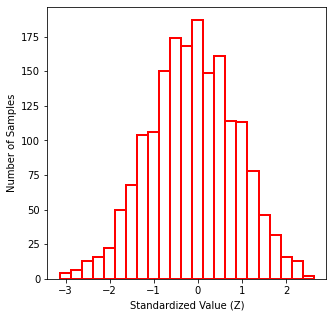
For full credit: write answers to the questions and then upload this document to your github along with notebook #1 (including any edits that you make).

1) Use the 2600-yearlong 1850 control run to calculate population statistics with constant forcing (in the absence of climate change). Find the population mean and population standard deviation for CESM1 global annual mean surface temperature. Normalize the data and again find the population mean and population standard deviation. Plot a histogram of the normalized data. Is the distribution Gaussian?

The population mean for the CESM1 global annual mean surface temperature is 287.11 K.

The population standard deviation for the CESM1 global annual mean surface temperature is 0.1 K.

The distribution is Gaussian.



2) Calculate global warming in the first ensemble member over a given time period defined by the *startyear* and *endyear* variables. Compare the warming in this first ensemble member with the 1850 control run statistics and assess if the warming is statistically significant. Use hypothesis testing and state the 5 steps. What is your null hypothesis? Try using a z-statistic (appropriate for N>30) and a t-statistic (appropriate for N<30). What is the probability that the warming in the first ensemble member occurred by chance? Change the startyear and endyear variables – When does global warming become statistically significant in the first ensemble member?

1. State the significance level: 5% ()

2. State the null hypothesis and the alternative: The null hypothesis is that the global mean annual temperature for the member 1 and for the population are the same.

3. State the statistics to be used.: t-statistic because N < 30 and one-tailed test because we expect warming.

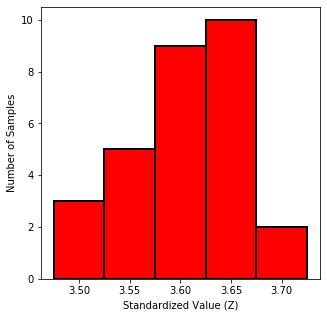
4. The null hypothesis is rejected if or if

5. Evaluate the statistics: Null hypothesis is rejected.

The probability that the warming in the first ensemble occurred by chance is 0% for both the t-statistic and the z-statistic.

Global warming becomes statistically significant in the first ensemble member from 1972 to 1982. Since the number of years is less than 30, we use the t-statistic. The t-statistic for 1972 to 1982 is 1.31, and for 1973 to 1983 is 2.38.

3) Many climate modeling centers run only a handful of ensemble members for climate change projections. Given that the CESM Large Ensemble has lots of members, you can calculate the warming over the 21st century and place confidence intervals in that warming by assessing the spread across ensemble members. Calculate confidence intervals using both a z-statistic and a t-statistic. How different are they? Plot a histogram of global warming in the ensemble members – Is a normal distribution a good approximation? Re-do your confidence interval analysis by assuming that you only had 6 ensemble members or 3 ensemble members. How many members do you need? Look at the difference between a 95% confidence interval and a 99% confidence interval.

The confidence interval obtained from the t-statistic are very similar to the confidence interval obtained from the z-statistic even when the number of ensemble members is around 5. The spread from the t-statistic is larger than for the z-statistic. The primary difference occurs for the 99% confidence limits. When the number of ensemble members is very low, then the distribution no longer resembles a normal distribution.  
 N = 30

N = 5