**ATOC5860 – Application Lab #1**

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

**in class Thursday January 20 and Tuesday January 25, 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 in** | **4.22 in** | **81 yrs** |
| **El Nino Years** | **15.29 in** | **4.0 in** | **16 yrs** |
| **La Nina Years** | **17.78 in** | **4.11 in** | **15 yrs** |

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

1. Our significance level is alpha = 0.05
2. Our null hypothesis is that there is no relationship between Nino3.4 SST Anomalies and Loveland Pass SWE. Our alternative hypothesis is that there is a physical relationship between Nino3.4 SST Anomalies and Loveland Pass SWE.
3. For this problem, we will be comparing the mean SWE during El Nino (La Nina) observed to the mean SWE obtained via 1000 bootstrapped each sampled El Nino N (La Nina N) times with a z statistic. This will be a two-tailed test because there are valid arguments for both sign changes.
4. The critical region will be defined below once the bootstrapping is completed.
5. The evaluation will be performed below.

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).
2. Quantify the likelihood of getting your value of mean SWE 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?

Chart, histogram

Description automatically generated

Mean 16.37 in

SD. 1.06 in

Min. 13.0 in

Max. 21.0 in

The probability that the difference in El Nino composite SWE from the population mean is due to chance is 30.7%. The probability that the difference in La Nina composite SWE from the population mean is due to chance is 18.19%.

Neither difference is sufficient to reject the null hypothesis. This data does not support the claim that either La Ninas or El Ninos result in a change in Loveland SWE.

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.

I am changing the El Nino (La Nina) Index to 0.5 (-0.5). This will result in the number of samples taken for each bootstrap to increase, giving me more data to work with and a smaller bootstrapped standard deviation.

Now N is 25 for El Nino and 27 for La Nina

Probability is now 35.75% for El Nino and 3.31 for La Nina. This isn’t really what I expected for La Nina. The data points between -0.5 and -1 must have a SWE measurement further away from the mean.

Interesting!

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!!

**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 standardized 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-year long 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. Standardize the data and again find the population mean and population standard deviation. Plot a histogram of the standardized data. Is the distribution Gaussian?

The mean prior to standardization is 287.11 K, and the standard deviation is 0.1 K. The mean and standard deviation after standardizing are 0.0 and 1.0, given with the definition of standardizing.

Chart, histogram

Description automatically generated

This distribution appears to be Gaussian.

1. 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 Our significance level is alpha = 0.05

1. Our null hypothesis is that the first ensemble mean temperature less than or equal to the control population mean temperature. Our alternative hypothesis is that the first ensemble mean temperature is significantly greater than the control population mean temperature.
2. For this problem we will compare the first ensemble mean temperature using both a z statistic and t statistic using multiple ranges for the first ensemble member data. I will first check 2020 to 2030 using a t test, then check 1920 to 2000 using a z test. These will both be one sided since we have prior reason to assume that warming has occurs.
3. For the t test, we are looking for t > 1.8331 to reject the null hypothesis. For the z test, we are looking for z > 1.64.
4. For 2020 to 2030, the t statistic calculated is 37.12 which is significantly larger than our critical value. This leads us to reject the null hypothesis. For 1920 to 1990, the z statistic calculated is 2.08. This also leads us to reject the null hypothesis.

Both analyses lead us to reject the null hypothesis. These findings suggest that there is a significant increase in the mean temperature from the control population to the first ensemble member.

1. 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 follow are confidence intervals for both z and t statistics for 30 ensemble members are given below

A screenshot of a computer

Description automatically generated with low confidence

The ranges are barely different for z and t tests, this is because the sample size is sufficiently large to make both tests similar.

Chart, bar chart, histogram

Description automatically generated

This distribution is somewhat normal, but I would say this is also skewed fairly to the right. I would hesitate to assume normality in this situation. I would want to perform a test to be sure.

1. ensemble members

A screenshot of a computer

Description automatically generated with low confidence Chart, bar chart, histogram

Description automatically generated

3 ensemble members

A screenshot of a computer

Description automatically generated with low confidence

Chart, bar chart, histogram

Description automatically generated

I would say 6 ensemble members is approximately normal, but going less than that is less. I still would want to perform a test for normalcy

In these smaller samples, the t ranges are distinctly larger than the z ranges.