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**	**4.22**	**81**
El Nino Years	**15.29**	**4.0**	**16**
La Nina Years	**17.78**	**4.11**	**15**

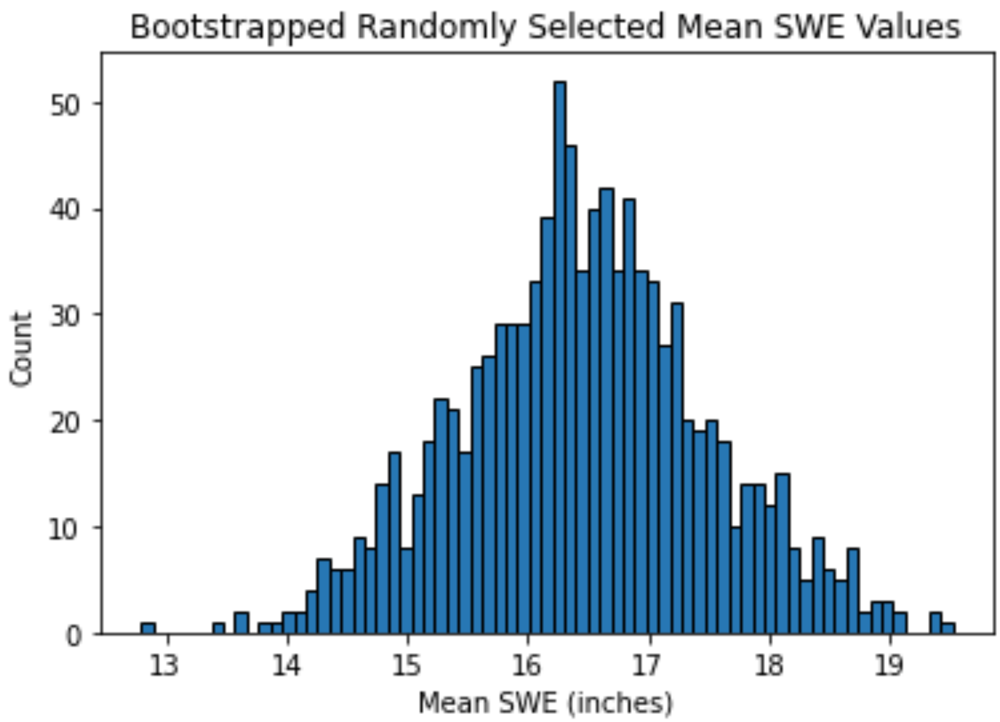
2) Use hypothesis testing to assess if the differences in snowpack are statistically significant. Write the 5 steps. Test your hypothesis using bootstrapping.

Five steps

1) Our significance level (alpha) is the 95% confidence level
2) State the null and alternative hypothesis:
    a. H_0: The means of the two samples (All Years versus El Nino or La Nina) are the same at the 95% confidence level)
    b. H_1: The means of the two samples (All Years versus El Nino or La Nina) are different at the 95% confidence level. El Nino year mean SWE will be less than all year mean SWE.
3) I will use bootstrapping to generate many randomly selected SWE means. I assumed that the Central Limit Theorem will apply as I am bootstrapping enough times. We then use the z-statistic to test our hypotheses.
4) We can use one-tailed z-test to see if we can reject the null hypothesis (need $Z < -1.65$ for El Nino)
5) For El Nino, the Z-statistic is -1.02, so we cannot reject the null hypothesis as it is not less than the test z-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.

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



Mean = 16.45
Standard Deviation = 1.03
Max = 12.79
Min = 19.52

b) 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?

The probability that differences between El Nino and all years occurred by chance is 13.15% (for a one-tailed test)

The probability that differences between the La Nina and all years occurred by chance is 9.86% (for a one-tailed test).

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.

Increasing the number of bootstraps does not change the the results of the experiment, but my group noticed that the percentage explained by chance seemed to converge with a larger N. Values of N at around 100,000 started to converge.

If the threshold for El Nino/La Nina is changed the ±1.5°C then the likelihood of the values occurring by chance increases! For El Nino, the one-sided test has a 27.44% chance. For La Nina, the one-tailed test has a 12.78% chance.

This suggests that changing the thresholds increased the odds of our results occurring by chance and that our results are not the most robust.

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!!

Running the other versions (t-test, the other bootstrapping approach), shows similar results, wherein we cannot reject the null hypothesis of the values being equal.

I found Vineel's bootstrapping method to be easier to understand.

<u>Notebook #2 – Statistical significance using z/t-tests</u>
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.

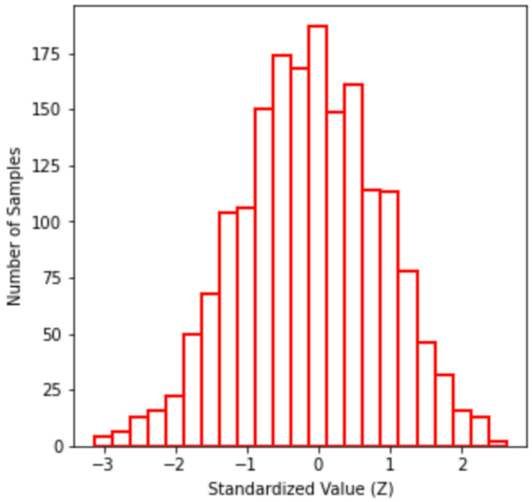
DATA and UNDERLYING SCIENCE:
You will be plotting *<u>munged</u>* climate model output from the Community Earth System Model (CESM) Large Ensemble Project. The Large Ensemble Project includes a 42-member ensemble of fully coupled climate model simulations for the period 1920-2100 (*note: only the original 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?

Population mean = 287.11 K
Population standard deviation = 0.1 K



This looks Gaussian to me!

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. 95% confidence
2. H_0: $T_{control} = T_{member\ 1}$
   H_1: $T_{control}$ ‡ $T_{member\ 1}$
3. For the situation when N>30 (assume normal distribution via CLT), we will use a z-statistic and when N<30 we will use a t-statistic
4. Our critical values are: $z_{crit} = 1.96$, $t_{crit} = 2.262$
5. Our z = 35.36, t = 37.12, so we can reject our null hypothesis

There is a zero percent chance that it is caused by chance!

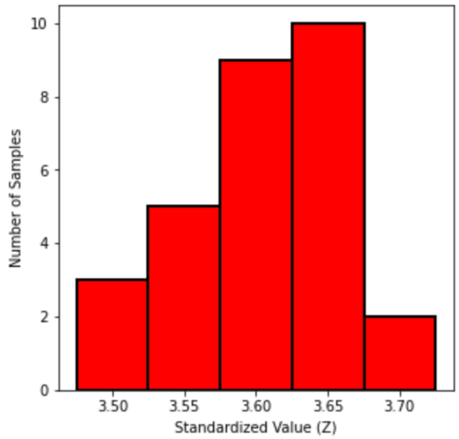
You start to see significance from 1980-1990!

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 intervals are very similar for all 4 renditions of creating confidence intervals! Ranging from 3.6-3.61 (low limit) to 3.66-3.67 (high limit).

This histogram shows that global warming spread is not a normal distribution!



With 6 ensemble assumption 95%: 3.59 - 3.68; 99%: 3.57 - 3.70

- This spread is similar to a larger sample

With 3 ensemble assumption 95%: 3.59 – 3.74; 99%: 3.49 - 3.83

- This spread is similar to a larger sample!

It seems that no matter how many members you have (or omit), we see a really strong warming signal!!