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).

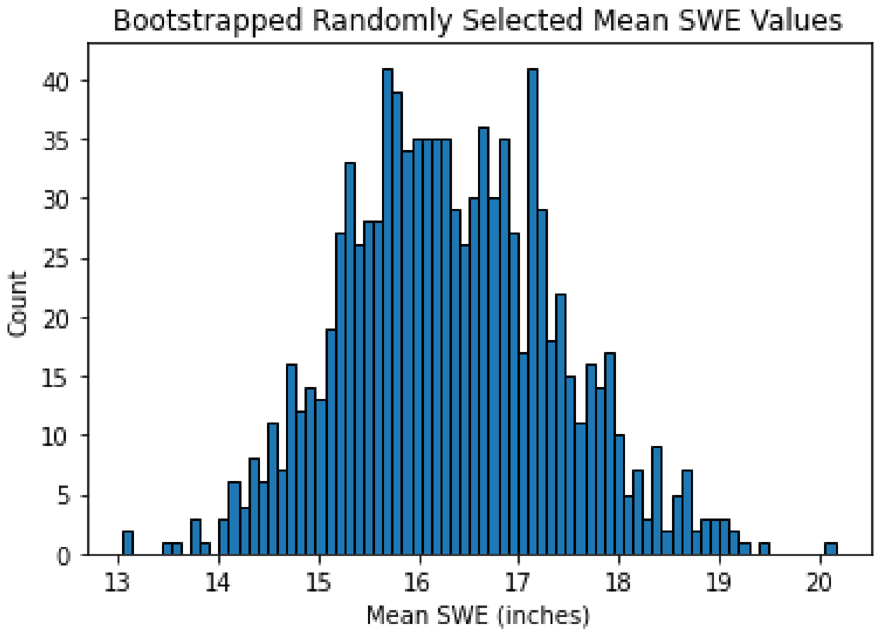
	Mean SWE (“)	Std. Dev. SWE (“)	N (# years)
All years	**16.33**	**4.22**	**81**
El Nino Years	**15.29**	**4.00**	**16**
La Nina Years	**17.78**	**4.11**	**15**

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

1. Significance Level: 95%
2. H_0: the mean SWE isn't lower during El Niño years
   H_1: the mean SWE is lower during El Niño years
   H_0: the mean SWE isn't higher during La Niña years
   H_1: the mean SWE is higher during La Niña years
3. The Bootstrap Method will be used to create a large number of samples for the distribution of mean SWE so that central limit theorem is assumed to be true. No other assumptions about the distribution are needed.
4. We can use a one-tailed test to see if the mean SWE is lower during El Niño years. ($z<-1.645$) and vice versa to see if mean SWE is higher during La Niña years ($z>1.645$). Both tests have a critical region of 5%.
5. The standardized mean SWE in El Niño years (-0.96) is not significantly lower than the mean SWE during all years, so we do not reject H_0.
The standardized mean SWE in La Niña years (1.38) is not significantly higher than the mean SWE during all years, so we do not reject H_0.

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).



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?

There is a 16.85% chance a mean SWE during El Niño years was as low as it is by chance.

There is an 8.41% chance a mean SWE during La Niña years was as high as it is by chance.

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.

My group noticed that N had to be increased to ~100,000 before the values became consistent between runs and for the histogram to appear to have an approximately normal distribution.

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!!

There still is not a significant difference between the means. I preferred the first method since I'm more familiar with the z-statistic and standardizing distributions, so I liked that more.

<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.

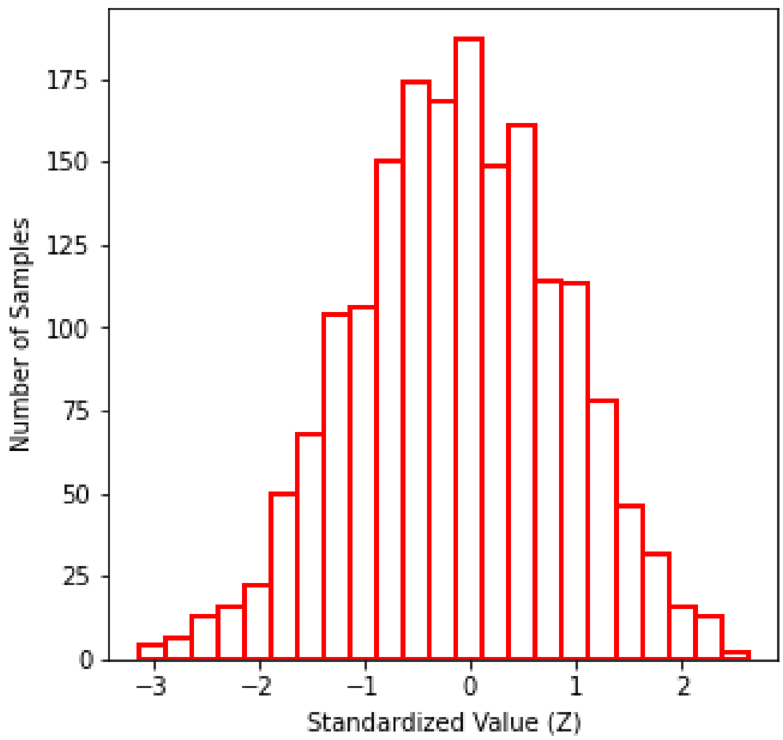
DATA and UNDERLYING SCIENCE:
You will be plotting <u>*munged*</u> climate model output from the Community Earth System Model (CESM) Large Ensemble Project. The Large Ensemble Project includes a 42-member ensemble of fully coupled climate model simulations for the period 1920-2100 (*note: only the original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

1850 Temperature mean = 287.11 K
1850 Temperature standard deviation = 0.1 K



The distribution appears to be Gaussian

2) Calculate global warming in the first ensemble member over a given time period defined by the start year and end year 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?

Given start year: 2020 and end year: 2030

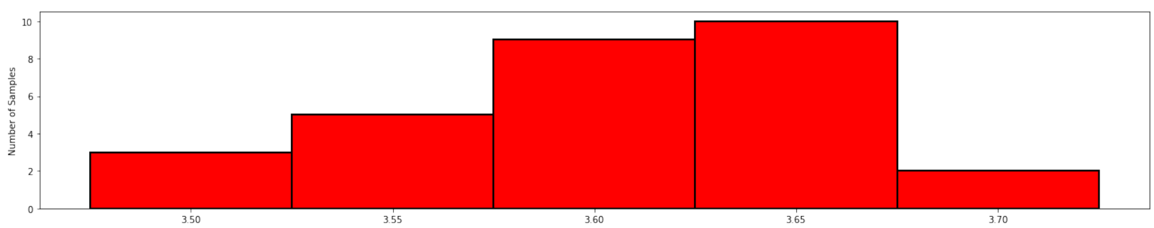
1. Significance Level: 95%
2. H_0: the mean ensemble temperature is higher than the 1850 control temperature
3. The distribution will be assumed to be normal
4. We can use a one-tailed test to see if the mean ensemble temperature is greater than the mean 1850 control run temperature (z>1.645) The test has a critical region of 5%.
5. The z statistic (35.36) is much larger than the critical z value (1.645), so cannot reject the null hypothesis.
The t statistic (37.12) is much larger than the critical z value (1.645), so cannot reject the null hypothesis.

It appears that there is approximately zero chance that the global rise in temperature is by chance. If you look at different decades, the null hypothesis cannot be rejected until 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.

95% confidence limits - t-statistic (3.61-3.66)
99% confidence limits - t-statistic (3.6-3.67)
95% confidence limits - z-statistic (3.61-3.66)
99% confidence limits - z-statistic (3.6-3.66)

The z and t confidence intervals are minimally different.



The data does not appear to be normal, it appears to be strongly skewed.
n=6 members
95% confidence limits - t-statistic (3.6-3.68)
99% confidence limits - t-statistic (3.57-3.71)
n=3 members
95% confidence limits - t-statistic (3.59-3.74)
99% confidence limits - t-statistic (3.49-3.83)
It seems that even with only n=3 the warming is clearly a strong and significant signal