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

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

State the significance level: Alpha = 0.05
State the null hypothesis and the alternative: H_0 = Mean April 1 SWE is the same during El Niño and all years. H_1 = Mean April 1 SWE is different during El Niño and all years.
State the statistic to be used and the assumptions that go into using it. Bootstrapping and two-tailed z-test. Z-test requires that the sample is large enough such that you know the population mean and standard deviation.
State the critical region: abs(z) > -1.96
Evaluate the statistic and state the conclusion. The statistic is z = -1.0, which means there is a 31.88% chance April 1 SWE between El Nino and all years purely due to chance. Thus there is not enough evidence to reject the null hypothesis.

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.

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

Mean = 16.38 inches
Standard deviation = 1.02 inches
Minimum = 12.68 inches
Maximum = 19.31 inches

Quantify the likelihood of getting your value of 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 30.81% chance that differences between the El Nino composite and all years occurred by chance.

There is a 17.2% chance that differences between the La Nina composite and all years occurred 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.

By changing the El Nino threshold to >0.5 degrees. The probability that the differences between the El Nino composite and all years occurred by chance increased slightly from 30.81 % to 32.29 %. So the conclusion is the same, thus lending some credibility to our results.

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!!

We do indeed get the same answer, although the precise confidence intervals and probabilities are not the same. I interpret this as a sign that whenever possible, it is best to draw conclusions from data with several different tests and techniques.

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-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. Normalize the data and again find the population mean and population standard deviation. Plot a histogram of the normalized data. Is the distribution Gaussian?

Population mean = 287.11 Kelvin
Population standard deviation = 0.1 Kelvin

Normalized population mean = 0.0
Normalized standard deviation = 1.0

Yes, the normalized distribution looks Gaussian/Normal!

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?

Significance level = 0.05

H_0: The temperatures in the control run and ensemble 1 are the same during the given time period.
H_1: The temperatures in the control run and ensemble 1 are different during the given time period.

We will use both the z and t statistics. Both require normal distributions, while the z statistic requires knowledge of the population mean and standard deviations.

Critical region: Probability that the larger temperature in ensemble 1 compared to the control simulation can be explained by chance is less than 5%.

For both the t-statistic and z-statistic the probability is 0%, thus we can reject the null hypothesis and claim the temperatures in the control run and ensemble 1 are different during the given time period.

The warming becomes significant in the 1980s.

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
3.657
99% confidence limits - t-statistic
3.601

3.665
95% confidence limits - z-statistic
3.611
3.656
99% confidence limits - z-statistic
3.604
3.663

The confidence intervals are very similar, probably because we have a relatively large number of ensemble members (N=30). But indeed, the t-statistic intervals are wider, which is to be expected.

The histogram of global warming is quassi-normal at best! So it retrospect it would probably be best to use a different method, i.e. boot strapping.

If I only use two ensemble members the 99% confidence interval still does not contain zero! So only 2 members are needed.