Gina Jozef
1/25/22

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/

Gina Jozef
1/25/22

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).

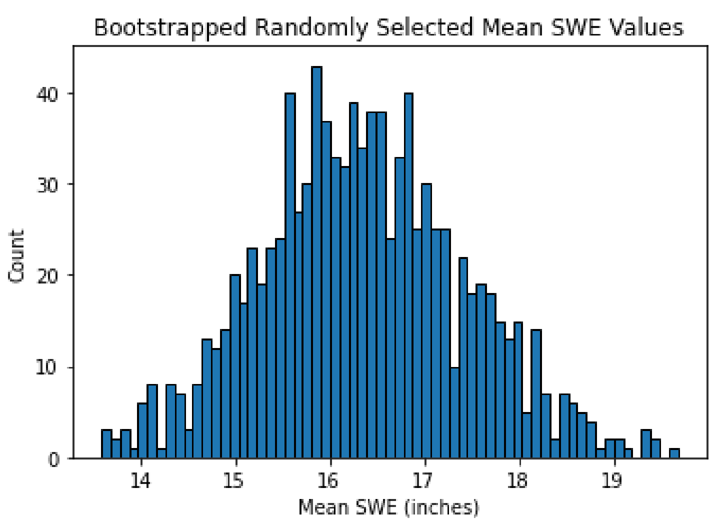
1) Composite Loveland Pass, Colorado snowpack. Fill out the following table showing the April 1 SWE in all years, in El Nino years (conditioned on Nino3.4 being 1 degree C warmer than average), and in La Nina years (condition on Nino3.4 being 1 degree C cooler than average).

	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.

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: 16.35 in
Stdev: 1.074 in
Min: 13.59 in
Max: 19.68 in

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?

Gina Jozef
1/25/22

1. State the significance level (α): 95% confidence ($\alpha = 0.05$).
2. State the null hypothesis, H_0, and alternative, H_1:
   El Nino:
   H_0: The snowfall (snow water equivalent) on April 1 was lower during El Nino years by chance.
   H_1: The snowfall (snow water equivalent) on April 1 was lower during El Nino years, not by chance.
   La Nina:
   H_0: The snowfall (snow water equivalent) on April 1 was higher during La Nina years by chance
   H_1: The snowfall (snow water equivalent) on April 1 was higher during La Nina years, not by chance.
3. We test the above null hypotheses, both using a one-sided z-test, since we specify which direction that we expect the difference in snowfall to be (lower during El Nino and higher during La Nina). We assume snowfall is standard normal.
4. The critical region: To reject the null hypothesis using a one-tailed z-statistic, the z-value must be less than (for El Nino) and greater than (for La Nina) z_c.
   El Nino
   This is a left-tailed test, so $z_c = -1.6449$ for $\alpha = 0.05$. Z-statistic must be less than z_c to reject the null hypothesis.
   La Nina
   This is a right-tailed test, so $z_c = 1.6449$ for $\alpha = 0.05$. Z-statistic must be greater than z_c to reject the null hypothesis.
5. El Nino
   Using the z-statistic, the z-value is -0.98, which is greater than the value of z_c. Calculating a probability value gives 16.68% probability that the snowfall on April 1 was lower during El Nino years by chance. Therefore, we cannot reject the null hypothesis.
   La Nina
   Using the z-statistic, the z-value is 1.33, which is less than the value of z_c. Calculating a probability value gives 7.84% probability that the snowfall on April 1 was higher during La Nina years by chance. Therefore, we cannot reject the null hypothesis.

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.

- Tried increasing the number of bootstraps from 1,000 to 100,000, keeping all other parameters the same as the original test. This does not change the results discussed above. The probability that differences in snowfall during El Nino and La Nina occurred by chance is still high enough to not reject the null hypotheses.

- Tried changing the temperature threshold that defines El Nino and La Nina to 0.5 degrees C, keeping all other parameters the same as the original test. This does not make a difference on whether we reject the null hypothesis for El Nino, but it does for La Nina. The z-statistic for La Nina is now 2.07 which is higher than z_c. Calculating a probability value gives 1.93% probability that the snowfall on April 1 was higher during La Nina years by chance. Therefore, we can reject the null hypothesis, since the probability is below 5%.
- Tried changing the significance level to 90% ($\alpha = 0.1$). This changes z_c to -1.28 for El Nino and 1.28 for La Nina. The z-statistic for El Nino is still not less than z_c in this case, but the z-statistic for El Nino is now greater than z_c in this case. Therefore, at the 90% significance level, we can reject the null hypothesis for La Nina.

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!!

- Using a t-test, the t-statistic is -1.04, while the t_c for a two-tailed t-test with 95% significance level and 15 degrees of freedom (since N=16) is ± 2.13, so we cannot reject the null hypothesis that the El Nino years have the same mean as the full record.
- Using Welch's test which compares two sample means, the t-statistic is 0.91, so we also cannot reject the null hypothesis that the El Nino years have the same mean as the full record.
- The approach provided by Vineel Yettella gives confidence intervals that include the value of zero difference between sample means, so by this method we also cannot reject the null hypotheses.

<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 *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?

Population mean: 287.11 K
Population standard deviation: 0.1 K

Standardized population mean: 0 K
Standardized population standard deviation: 1 K

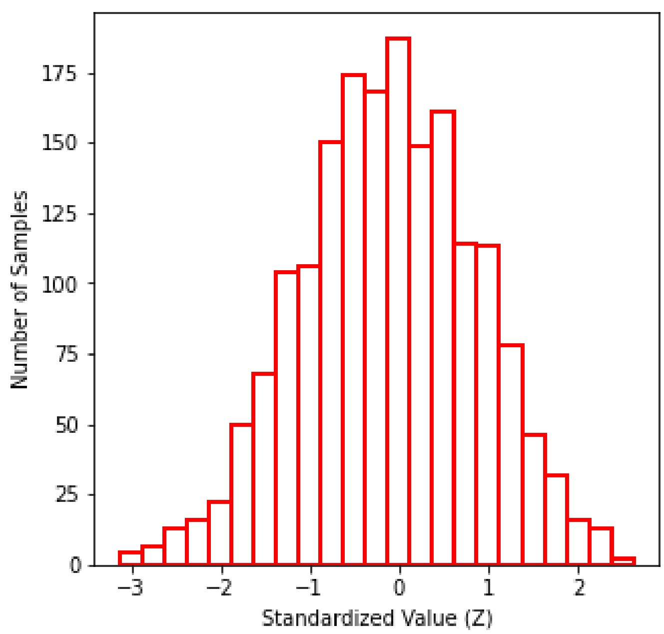


Figure 1: Frequency of standardized value of 1850 global mean surface temperature.

The above distribution is Gaussian.

Gina Jozef
1/25/22

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?

<u>5 steps:</u>

1. State the significance level (α): 95% confidence ($\alpha = 0.05$).
2. State the null hypothesis, H_0, and alternative, H_1:
   H_0: the sample mean of annual global mean surface temperatures in the first ensemble member between 2020 and 2030 is equal to the population mean.
   H_1: the sample mean of annual global mean surface temperatures in the first ensemble member between 2020 and 2030 is significantly different than the population mean.
3. Testing this using both a two-sided t-test and z-test. We assume the global mean annual temperature is standard normal.
4. The critical region: To reject the null hypothesis using the t-statistic, the t-value must be outside of than $\pm t_c$, which in this case, with 9 degrees of freedom, is ± 2.262. To reject the null hypothesis using the z-statistic, the z-value must be outside of $\pm z_c$ which is ± 1.96 for $\alpha = 0.05$.
5. Using the t-statistic, the t-value is 37.12, which is much greater than the value of t_c. Using the z-statistic, the z-value is 35.36, which is much greater than the value of z_c. Calculating a probability value for both tests gives 0% probability that the warming in the first ensemble member occurred by chance. Therefore, we reject the null hypothesis.

   To test when global warming became statistically significant, I changed the startyear to 1920, which was the beginning time of the CESM large ensemble data, and also before global warming became significant. I then changed the endyear to different years increasing from 1920, until the probability that the warming in the first ensemble occurred by chance went below 5%, which indicates that global warming became statistically significant. For both the t-test and z-test, this occurs when endyear is 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.

Confidence intervals of warming over the 21st century using all 30 ensemble members:
t-statistic:
95% confidence limits: 3.61 – 3.66 K
99% confidence limits: 3.6 – 3.67 K
z-statistic:
95% confidence limits: 3.61 – 3.66 K
99% confidence limits: 3.6 – 3.66 K

These are very similar, likely because N = 30, which is the point where the t and z statistic will give almost the same results.

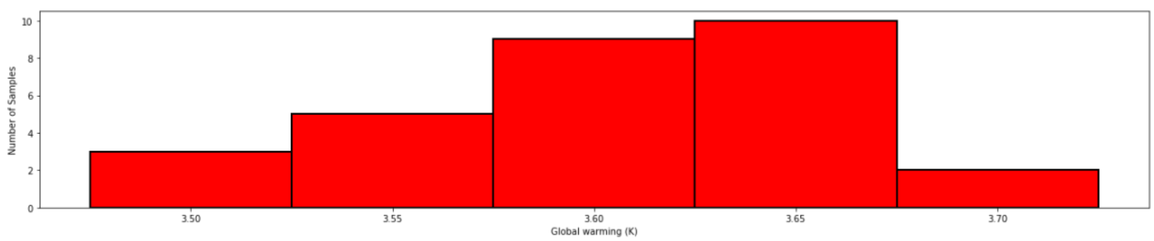


Figure 2: Histogram of global warming in the ensemble members.

The above histogram shows essentially a normal distribution, so normal distribution is a good approximation.

Confidence intervals of warming over the 21st century using all 6 ensemble members:
t-statistic:
95% confidence limits: 3.59 – 3.68 K
99% confidence limits: 3.57 – 3.71 K
z-statistic:
95% confidence limits: 3.61 – 3.67 K
99% confidence limits: 3.6 – 3.68 K

Confidence intervals of warming over the 21st century using all 3 ensemble members:
t-statistic:
95% confidence limits: 3.59 – 3.74 K
99% confidence limits: 3.49 – 3.83 K
z-statistic:
95% confidence limits: 3.63 – 3.69 K
99% confidence limits: 3.63 – 3.7 K

To show that global warming over the 21st century is significant, you only need 1 ensemble member for the z-test, and 2 for the t-test (since the t-test doesn't work when N=1). This gives high confidence that global warming over the 21st century is occurring.