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 (in.)	**Std. Dev. SWE (in.)**	**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.

El Nino:

1) State the significance level (α)
   The significance level I have chosen is 5% (or 95% confidence).
2) State the null hypothesis H_0 and the alternative H_1
   mu_bootstrapped_SWE _samples = mu_SWE_el_nino
   mu_bootstrapped_SWE _samples ≠ mu_SWE_el_nino
3) State the statistic to be used, and the assumptions required to use it
   The statistic to be used is bootstrapping. By using the bootstrapping method, I am able to use the z-test and will not assume a normal distribution.
4) State the critical region
   I will use a two-tailed z-test. Thus, to reject the null hypothesis I must have $|t| > 1.96$.
5) Evaluate the statistic and state the conclusion
   $|z| = -0.98$
   z is not greater than -1.96 so we cannot reject the null hypothesis.

La Nina:

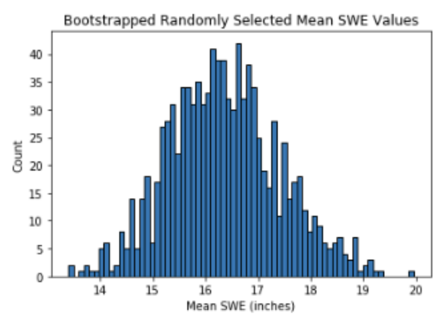
1) State the significance level (α)
   The significance level I have chosen is 5% (or 95% confidence).
2) State the null hypothesis H_0 and the alternative H_1
   mu_bootstrapped_SWE _samples = mu_SWE_la_nina
   mu_bootstrapped_SWE _samples ≠ mu_SWE_la_nina
3) State the statistic to be used, and the assumptions required to use it

The statistic to be used is bootstrapping. By using the bootstrapping method, I am able to use the z-test and will not assume a normal distribution.

4) State the critical region
    a. I will use a two-tailed z-test. Thus, to reject the null hypothesis I must have |t| > 1.96.
5) Evaluate the statistic and state the conclusion
    a. The probability is 15.67%, which falls within the 95% confidence interval. Therefore, the average SWE in 15 La Nina years between 1963-1940 is not significantly different from the average SWE for all years 1936-1940 combined.

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).



The distribution is close to normal. The mean is 16.35 inches, standard deviation is 1.05 inches, the minimum value is 13.39 inches and the maximum value is 19.95 inches.

b) Quantify the likelihood of getting your value 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 is 32.44%, which falls within the 95% confidence interval. Therefore, the average SWE in 16 El Nino years between 1963-1940 is not significantly different from the average SWE for all years 1936-1940 combined.

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.

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!!

1) State the significance level (alpha symbol (1-signifcant)) 5%
2) State the null hypothesis H_0 and the alternative H_1

    mu_bootstrapped_SWE _samples - mu_SWE_el_nino = 0

    mu_bootstrapped_SWE _samples- mu_SWE_el_nino ≠ 0
3) State the statistic to be used, and the assumptions required to use it

    Bootstrapping will be used. We assume there is a normal distribution.
4) State the critical region

    We are really only looking for a zero in the confidence interval or not.
5) Evaluate the statistic and state the conclusion

    The confidence interval contains zero, so we fail to reject the null hypothesis that the difference in means is equal to zero.

<u>Notebook #2 – Statistical significance using z/t-tests</u>
ATOC7500_applicationlab1_ztest_ttest.ipynb

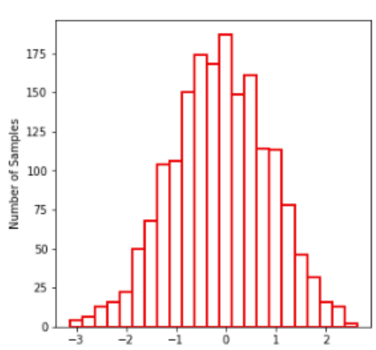
LEARNING GOALS:
1) Use an ipython notebook to read in a netcdf file, make line plots and histograms, and calculate statistics
2) Calculate statistical significance of the changes in a normalized mean using a z-statistic and a t-statistic
3) Calculate confidence intervals for model-projected global warming using z-statistic and t-statistic.

DATA and UNDERLYING SCIENCE:
You will be plotting *<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. Normalize the data and again find the population mean and population standard deviation. Plot a histogram of the *<u>standardized</u>* data. Is the distribution Gaussian?



This distribution appears to be Gaussian (normally distributed).

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?

z-statistic (N>30) for 2020-2030

1) State the significance level (α)
   The significance level I have chosen is 5% (or 95% confidence).
2) State the null hypothesis H_0 and the alternative H_1
   mu_10yr_sample = mu_control_run
   mu_10yr_sample ≠ mu_control_run
3) State the statistic to be used, and the assumptions required to use it
   The statistic to be used is a z-test statistic.
4) State the critical region
   I will use a one-tailed z-test. Thus, to reject the null hypothesis I must have |t| >= 1.65.
5) Evaluate the statistic and state the conclusion
   |z| = 35.36
   z is greater than 1.65 so we can reject the null hypothesis.

There is zero probability that the warming in the first ensemble member occurred by chance.

t-statistic (N<30) for 2020-2030

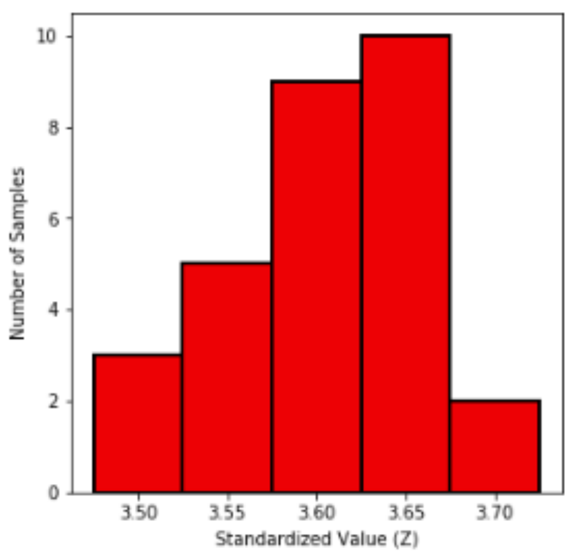
1) State the significance level (α)
   The significance level I have chosen is 5% (or 95% confidence).
2) State the null hypothesis H_0 and the alternative H_1
   mu_10yr_sample = mu_control_run
   mu_10yr_sample ≠ mu_control_run
3) State the statistic to be used, and the assumptions required to use it
   The statistic to be used is a t-test statistic.
4) State the critical region
   I will use a two-tailed z-test. Thus, to reject the null hypothesis I must have |t| >= 1.833.
5) Evaluate the statistic and state the conclusion
   |t| = 37.12

t is greater than 1.833 so we can reject the null hypothesis.
There is zero probability that the warming in the first ensemble member occurred by chance.

Global warming becomes statistically significant when I first change the startyear and endyear variables to 1970 and 1980 for the first ensemble member, and then change the startyear and endyear variables to 1980 and 1990. Between 1970 and 1980, the t-statistic is 0.63 so we cannot reject the null hypothesis. Between 1980 and 1990, the t-statistic is 4.32 and we can reject the null hypothesis.

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.

z-statistic 95% confidence intervals for the spread of CESM Large Ensemble (CESM-LE) members are 0.05 degrees C. t-statistic 95% confidence intervals for the spread of CESM Large Ensemble (CESM-LE) members are 0.07 degrees C. They are different by 0.02 degrees C.



Using 30 CESM-LE members, a normal distribution is not a good approximation. Experimenting with 6 CESM-LE members presents a normal distribution. Using 3 CESM-LE members does not.