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**	**4.0**	**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.

1. State the significance level: $\alpha = 0.05$
2. State the null hypothesis:

   $\mathcal{H}_0$: $\mu_{\text{el nino}} = \mu_{\text{bootstrapped samples}}$

   $\mathcal{H}_1$: $\mu_{\text{el nino}} \neq \mu_{\text{bootstrapped samples}}$
3. For this problem we will use bootstrapping and then a two-sided z-statistic.
4. For this problem we do not know what to expect with regards to how Nino3.4 affects Loveland Pass SWE. Thus to reject the null hypothesis $|z| \geq 1.96$
5. Evaluate statistic and state conclusion:

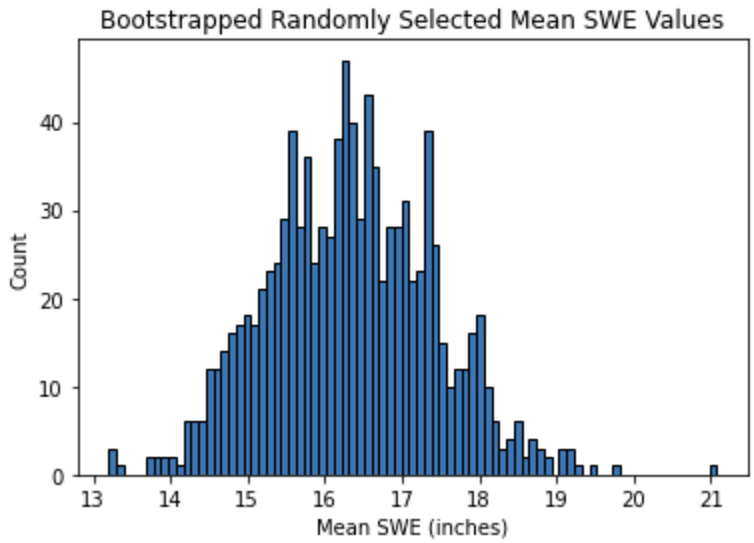
$$z = \frac{\mu_{sample} - \mu_{population}}{\frac{\sigma_{population}}{\sqrt{N}}} = -0.96$$

$|z|$ is not greater than -1.96 so we fail to reject the null hypothesis.

If we do a one-sided z-test then z = -1.06, thus we would still fail 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.

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



mean = 16.38 in
standard deviation = 1.02 in
min = 12.95 in
max = 19.27 in

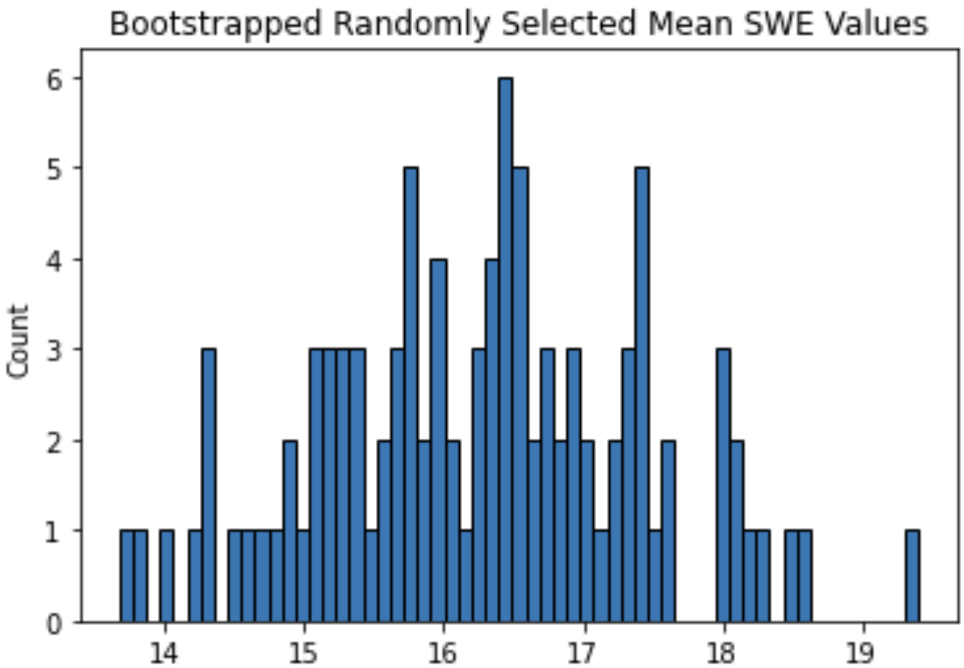
b) Quantify the likelihood of getting your value of by chance alone using percentiles of this bootstrapped distribution. What is the probability that differences between the El Nino composite and all years occurred by chance? What is the probability that differences between the La Nina composite and all years occurred by chance?

The probability that the lower SWE during El Nino years is by random chance is 33.64 %

The probability that the lower SWE during El Nino years is by random chance is 16.99 %

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.

If we bootstrap fewer times (say 100 instead of 1000) then the distribution of means gets farther from a normal distribution (see below). However, bootstrapping fewer times does not change the final conclusion as we still fail to reject the null hypothesis that SWE during el nino/la nina years is significantly different from normal.



If we increase the temperature threshold for an el Nino/la Nina year to 2 deg temperature anomaly then fewer years are classified as el Nino or la Nina years. After bootstrapping 1000 times the z-statistic for el Nino years decreases (to -0.14) while the z-statistic for la Nina years increases (to 1.2). Regardless, the conclusion remains the same that we fail to reject the null hypothesis.

It seems that regardless of the parameters we have changed the conclusion remains the same in that SWE during el Nino and la Nina years is not significantly different from normal (unless we were to increase the significance level substantially). Thus, our conclusion seems fairly 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!!

For el Nino years:
With a t-test we get a t-statistic of 0.91 and a p-value of 0.367, thus we fail to reject the null hypothesis.

With Vineel's Yettalla's method we get a 95% confidence interval of -1.02 to 3.14 for the difference between el Nino mean SWE and the bootstrap mean. Because zero lies in the confidence interval we fail to reject the null hypothesis.

<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.

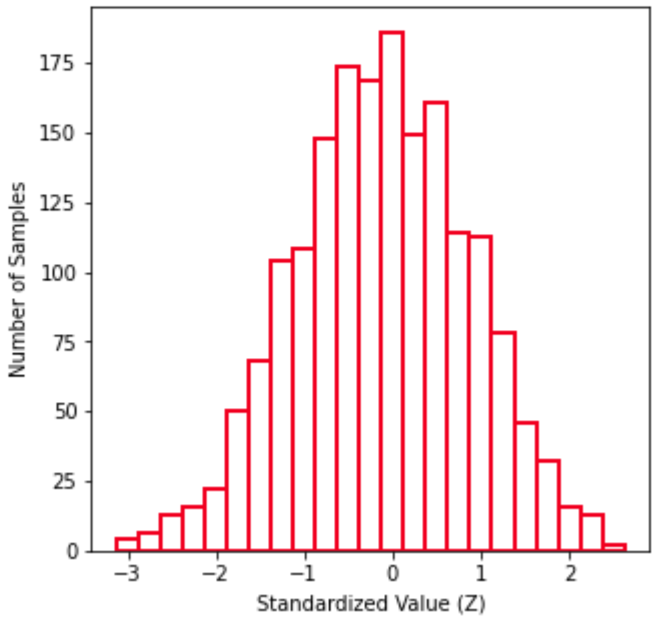
DATA and UNDERLYING SCIENCE:
You will be plotting <u>*munged*</u> climate model output from the Community Earth System Model (CESM) Large Ensemble Project. The Large Ensemble Project includes a 42-member ensemble of fully coupled climate model simulations for the period 1920-2100 (*note: only the original 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. Standardize the data and again find the population mean and population standard deviation. Plot a histogram of the standardized data. Is the distribution Gaussian?

mean =	13.96 deg C
standard deviation =	0.1 deg C



The mean of the standardized data is 0 and the standard deviation is 1. The distribution is Gaussian.

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. State the significance level: $\alpha = 0.05$
2. State the null hypothesis:
   $\mathcal{H}_0$: $\mu_{sample} = 0$
   $\mathcal{H}_1$: $\mu_{sample} \neq 0$
3. For this problem we will first use a one-sided z-test and then we will use a one-sided t-test. We are using a one-sided test because we expect that the sample mean is larger than the control mean due to global warming.
4. To reject the null hypothesis $z > 1.65$ or for t-test: $t > 1.833$
5. We get a z-statistic of 35.36 and a t statistic of 37.12. Thus, we reject the null hypothesis

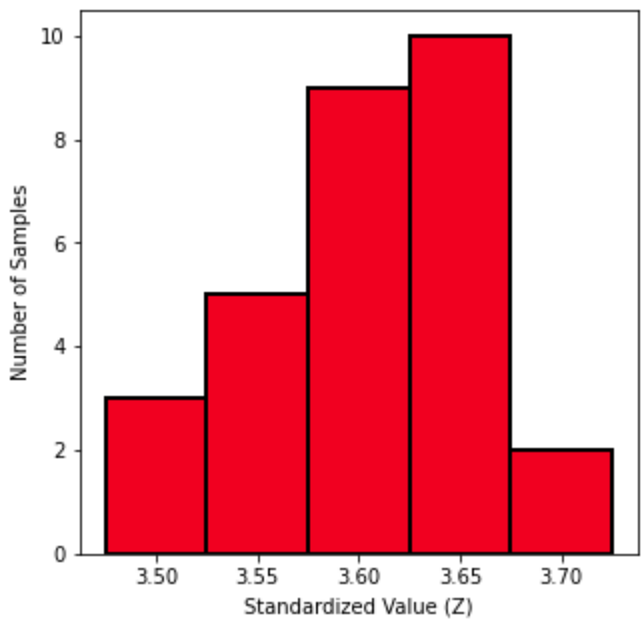
Using the years 1970-1980 we get a t-statistic of 0.63, in which case we fail to reject the null hypothesis. However, using the years 1980-1990, we get a t-statistic of 4.32 in which case we reject the null hypothesis.

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.

95% confidence interval using z-statistic: 3.61 - 3.66
95% confidence interval using t-statistic: 3.61 - 3.66
99% confidence interval using z-statistic: 3.6 - 3.67
99% confidence interval using t-statistic: 3.6 - 3.66
The confidence intervals for z-statistics and t-statistics are very similar (identical for the 95% confidence interval).

Histogram of global warming in the ensemble members:



To me, a normal distribution does not appear to be a super great approximation. Perhaps it would help to include more ensemble members.

For 6 ensemble members:
- 95% confidence interval using t-statistic: 3.595-3.682
- 99% confidence interval using t-statistic: 3.511 - 3.766

For 3 ensemble members:
- 95% confidence interval using t-statistic: 3.587-3.736
- 99% confidence interval using t-statistic: 3.582 - 3.741