1a. We want to know the probability of actually being negative given the test's reliability of saying you are negative, P(N|T). To put this simply, this is the fraction of the probability that the test is correct divided by the combined probability of all possible outcomes:

$$P(N|T) \ = \ \frac{P(T|N)\,P(N)}{P(T|N)\,(P(N) + P(T|P)(P(P)}$$

The equation above is adjusted from Barnes, equation 23

If I were to explain this to a friend, I would talk about how the COVID test accuracy at face value can be deceiving. This actually depends on how much of the population is generally negative and or positive. In an environment where half or more of the population is more likely to be positive, the accuracy of the test matters more because the probability of a test showing someone is negative and they are truly negative is highly dependent on all possible outcomes. If the ratio of COVID Positive to COVID Negative is even or higher, then there is an increased probability of a false negative. Likewise, if this ratio is lower, results can be trusted to a higher confidence.

1b. Testing whether a sample mean is significantly different than zero at the 95% or 99% confidence interval can depend on the number of observations. If the number of observations is less than 30 (N<30), then a t-test is more appropriate to apply. Otherwise a z-test is more appropriate. It is worth noting that for larger degrees of freedom (defined as N-1), a t-test and z-test will eventually converge to similar distributions. Before we apply a test we need to outline the five hypothesis testing steps:

1. State the significance level, α. At the 95% confidence interval, α=0.05. At the 99% confidence interval, α=0.01. Since one of the tasks uses N=15, we must use a t-test for that hypothesis but we can use either a t-test or z-test for N=1000.
2. State the null and alternative hypotheses:
    a. H0: The sample mean **does not** differ from the population mean.
    b. H1: The sample mean **does** differ from the population mean.
3. State the statistic to be used and the assumptions required to use it.
    a. We must assume there is a normal distribution, and we must pull 15 or 1000 independent observations depending on the experiment.
4. State the critical region.
    a. Critical values can be retrieved from the t-test table or with the *scipy.stats.t.ppf* function. From the table, the t-critical value is 2.145 at the 95% confidence level and 2.977 at the 99% confidence level for 15 degrees of freedom. For N=1000 independent observations, the t-critical value is 1.962 at the 95% confidence level and 2.581 for the 99% confidence level.
5. Evaluate the statistic and state the conclusion.
    a. For a two-tailed t-test if the sample mean for either experiment differs from the population mean and that corresponding t-statistic is less than the t-critical value then we can reject the null hypothesis.

NOTE: For problems 1c-e, I chose to use an image pair from my personal research. The data are synthetic images produced by our group member, Dr Hong Chen, to assist with my parallax cloud detection algorithm with a known ground truth. If you choose to run this on your own computer, you will need to install the h5py library into the environment.

1c. I compared the radiance of images with a viewing zenith angle (VZA) at 0 and 30 degrees in a defined area that was confirmed to have clouds using a ground truth mask. The goal was to see if I could find a sample mean difference between the two images taken at the same location but with differing VZA.

My five steps in hypothesis testing were:

1. Conduct a two-tailed t-test at the 95% significance level (α=0.05) by applying a two-tailed t-test for 1000 bootstrapped cloud areas, indicated as **sample 1** and **sample 2**.
2. State the hypotheses:
    a. H0: Sample mean 1 **does not diffe**r from sample mean 2
    b. H1: Sample mean 1 **does differ** from sample mean 2.
3. Using the bootstrap approach, we can safely assume (but will verify) that the samples 1 and 2 are normally distributed. We will then apply the Welch's T-Test to determine if the sample means are significantly different from each other.
4. The critical value at the 95% confidence is 1.962.
5. Evaluate the statistics through Welch's T-Test and state the conclusion.

When plotting the distributions, the radiances were abnormally distributed (fig 1). To provide a more appropriate distribution I chose to bootstrap 1000 samples, 50 times. While I'm not sure if it is appropriate to analyze radiance distributions in this manner, it had the effect I wanted to accomplish the problem. Bootstrapping produced more normal distributions to analyze (fig 2).

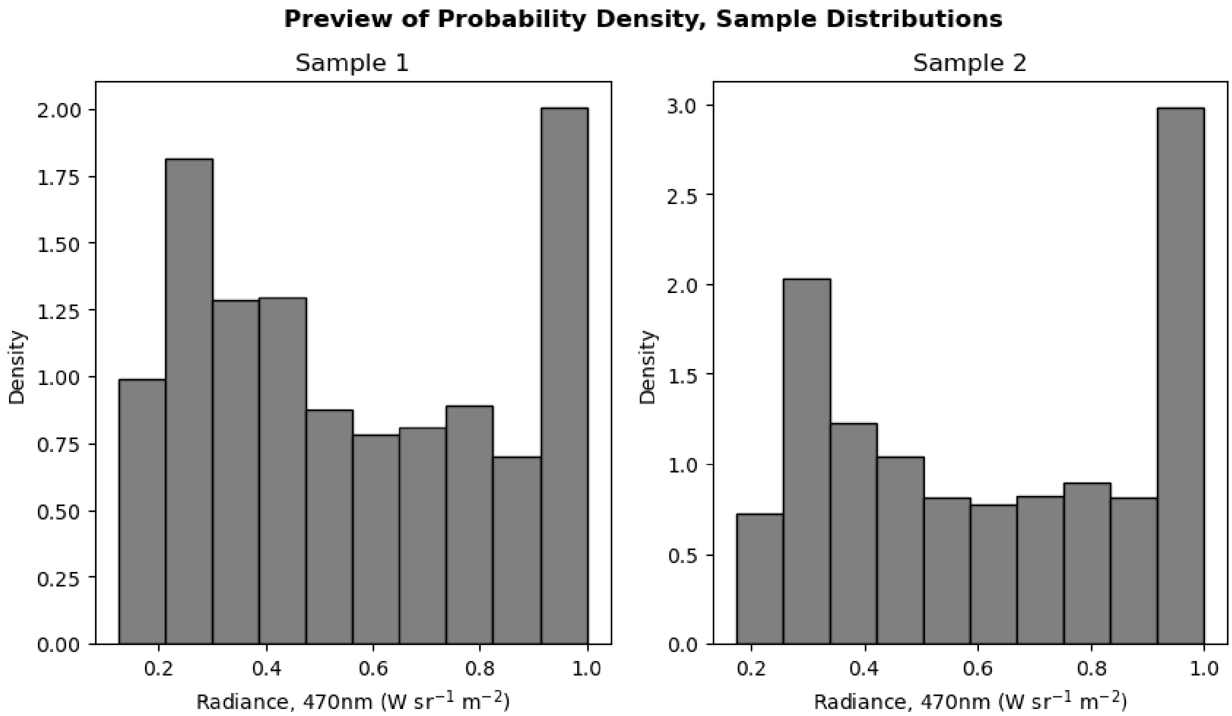

Fig 1. Abnormal distribution of samples 1 and 2, before the bootstrap method.

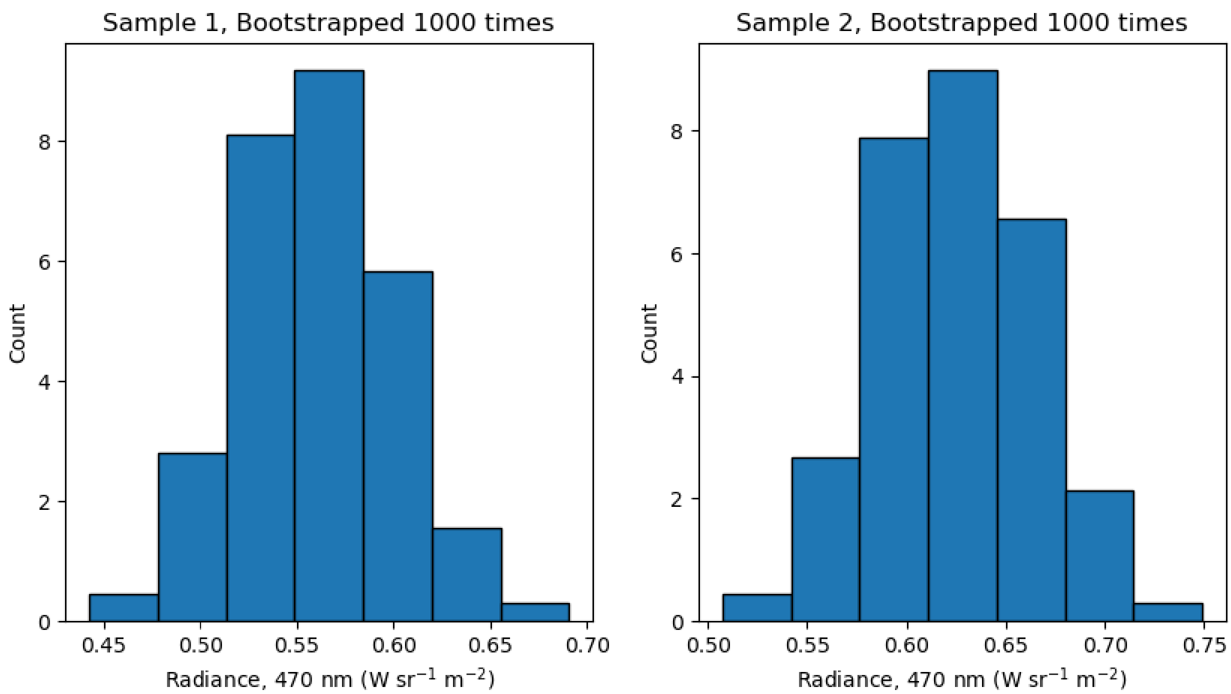

Fig 2. Normal distribution of samples 1 and 2, after the bootstrap method.

After calculating the variance and standard deviations for both samples, I applied the Welch's T-Test through the *scipy.stats.ttest_ind* function due to the slightly different variances. This function was useful because I could still apply this function by setting the *equal_var* keyword argument (*kwarg) to true or false. The magnitude of difference between these variances were very small, but it was interesting to investigate how applying Welch's T-Test assuming unequal and equal variances differed in p-values, but ultimately came to the same conclusion: the t-statistic in both cases were near -36.85 with a p-value much less than 0.05. Because of this p-value I was able to reject the null hypothesis.

1d) Using this same dataset, I then calculated the 95% confidence intervals for sample 1. The intervals were calculated with the *np.percentile* function to determine the appropriate results for values below the 2.5th and 97.5th percentile. This yielded a confidence interval of $0.487 < X < 0.641$.

1e) With the bootstrapped sample 1 and sample 2 data, I compared the ratio between the two standard deviations and their corresponding variance. Although the variance was highly similar, I chose to use Barnes Equation 121, where s is standard deviation and σ is the variance:

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

The result of this expression with my calculated standard deviation and variances was 0.967.
2a) The average hourly pressure in 2014 was 846.33 hPa. The average pressure when it rained was 847.3 hPa.

2b) After normalizing the data, the 95% confidence interval for a two-tailed t-test was $-0.1<X<0.1$. The normalized average pressure when it was raining was 0.1243 with a 0.74% chance that this sample mean occurred by chance, indicating that it is statistically significant.

2c) The bootstrap method provided a 95% confidence interval of $-0.227 < X < =0.024$. The resampled bootstrap mean, was 0.14, which aligned with the previous conclusion that the pressure is anomalously high during precipitation events.