Conducting T-Tests

Introduction

Just as you previously used the t distribution to provide confidence intervals for estimating the population mean, you can also use similar methods to test whether two populations are different, statistically speaking. To do this, you can use a t-test.

Objectives

You will be able to:

- Compare when you would use one sample vs. two sample t-tests
- Perform a one sample t-test and make conclusions about an experiment based on the results

Hypothesis testing using the T-distribution

In frequentist hypothesis testing, you construct a test statistic from the measured data and use the value of that statistic to decide whether to accept or reject the null hypothesis. The test statistic is a lower-dimensional summary of the data but still maintains the discriminatory power necessary to make the decision whether or not to reject the null hypothesis.

t-test

t-tests (also called Student's t-test) are very practical hypothesis tests that can be employed to compare two averages (means) to assess if they are different from each other. You should run a t-test when you either:

- Don't know the population standard deviation
- You have a small sample size

Like a z-test, the t-test also tells you how significant the differences are i.e. it lets you know if those differences could have happened by chance. In this lesson, you will get an introduction to t-tests, in particular, the 1-sample t-test. There are additional kinds of t-tests including the 2-sample t-test and paired t-test. This lesson will show you the mathematical calculations behind a 1-sample t-test as well as how to perform a t-test in Python using NumPy and SciPy.

Detailed descriptions of hypothesis testing with t-tests can be found here and here

One Sample t-test

The 1-sample t-test is a statistical procedure used to determine whether a sample of observations could have been generated by a process with a specific mean. The one sample t-test compares the mean of your sample data to a known value. For example, you might want to know how your sample mean compares to the population mean. Here is a quick example of a scenario where a 1-sample t-test could be applied.

Suppose you are interested in determining whether a bakery production line produces cakes with a weight of exactly 2 pounds. To test this hypothesis, you could collect a sample of cakes from the production line, measure their weights, and compare the sample with a value of 2 using a one-sample t-test.

Two Sample t-tests

The two-sample t-test is used to determine if two population means are equal. The main types of two-sampled t-tests are paired and independent tests. Paired tests are useful for determining how different a sample is affected by a certain treatment. In other words, the individual items/people in the sample will **remain the same** and researchers are comparing how they change after treatment. Here is an example of a scenario where a two-sample paired t-test could be applied:

The US Olympic weightlifting team is trying out a new workout techniques to in an attempt to improve everyone's powerlifting abilities. Did the program have an effect at a 95% significance level?

Because we are looking at how specific individuals were affected by a treatment, we would use the paired t-test.

Independent two-sample t-tests are for when we are comparing two different, unrelated samples to one another. Unlike paired t-tests, we are not taking paired differences because there is no way to pair two unrelated samples! Here is an example of a scenario where a two-sample independent t-test could be applied:

Agricultural scientists are trying to compare the difference in soybean yields in two different counties of Mississippi.

You will learn more about the specifics of two sample t-tests in future lessons, but this lesson will focus on executing a one sample t-test.

Assumptions for performing t-tests

When performing various kinds of t-tests, you assume that the sample observations have numeric and continuous values. You also assume that the sample observations are independent from each other (that is, that you have a simple random sample) and that the samples have been drawn from normal distributions. You can visually inspect the distribution of your sample using a histogram, for example.

In the case of unpaired two-sample t-tests, you also assume that the populations the samples have been drawn from have the same variance. For paired two-sample t-tests, you assume that the *difference* between the two sets of samples are normally distributed.

Regardless of the type of t-test you are performing, there are 5 main steps to executing them:

1) Set up null and alternative hypotheses

2) Choose a significance level

3) Calculate the test statistic

4) Determine the critical or p-value (find the rejection region)

5) Compare t-value with critical t-value to accept or reject the Null hypothesis.

Now, you're going to go through these 5 steps in more detail to complete a t-test.

Let's begin with a sample experiment:

Sample question:

> *"Acme Ltd. wants to improve sales performance. Past sales data indicate that the average sale was 100 dollars per transaction. After training the sales force, recent sales data (from a random sample of 25 salesmen) is shown below:"*

```
[122.09, 100.64, 125.77, 120.32, 118.25,
  96.47, 111.4 ,  80.66, 110.77, 111.14,
 102.9, 114.54,  88.09,  98.59,  87.07,
 110.43, 101.9 , 123.89,  97.03, 116.23,
 108.3, 112.82, 119.57, 131.38, 128.39]
```

Did the training work?

Before completing the hypothesis test, let's calculate some summary statistics to see if the mean of the sample differed a substantial amount from the population. After, you can check to ensure that the data is relatively normal.

- **The population mean (μ):** Given as 100 (from past data).
- **The sample mean ($\bar{x}$):** Calculate from the sample data
- **The sample standard deviation (s):** Calculate from sample data
- **Number of observations(n):** 25 as given in the question. This can also be calculated from the sample data.
- **Degrees of Freedom(df):** Calculate from the sample as df = total no. of observations - 1

In [1]:

```python
## Import the packages
import numpy as np
from scipy import stats
import math

# For visualizing distributions - optional
import seaborn as sns
import matplotlib.pyplot as plt
sample = np.array([122.09, 100.64, 125.77, 120.32, 118.25,  96.47, 111.4 , 80.66,
       110.77, 111.14, 102.9 , 114.54,  88.09,  98.59,  87.07, 110.43,
       101.9 , 123.89,  97.03, 116.23, 108.3 , 112.82, 119.57, 131.38,
       128.39])
```

```
# Population mean (μ)
mu = 100

# Sample mean (x̄) using NumPy mean()
x_bar = sample.mean()

# Sample Stadrad Deviation (sigma) using Numpy
sigma = np.std(sample,ddof=1)

# Sample size (n)
n = len(sample)

# Degrees of Freedom
df = n-1

# Difference in sample mean
diff = x_bar - mu


# Print the findings
print ('The sample contains', n, 'observations, having a mean of', x_bar, "and a standard deviation (si
gma) = ", sigma,
       ", with", df, 'degrees of freedom. The difference between sample and population means is:', diff
)

# The sample contains 25 observations, having a mean of 109.5456
# and a standard deviation (sigma) =  13.069276668584225 ,
# with 24 degrees of freedom.
# The difference between sample and population means is: 9.54
```

The sample contains 25 observations, having a mean of 109.5456 and a standard deviation (sigma) = 13.3
38774643871902 , with 24 degrees of freedom. The difference between sample and population means is: 9.5
45599999999993

The sample mean is $9.54 per sale higher than the population mean, which indicates that, at least superficially, the training program appears to have improved performance.

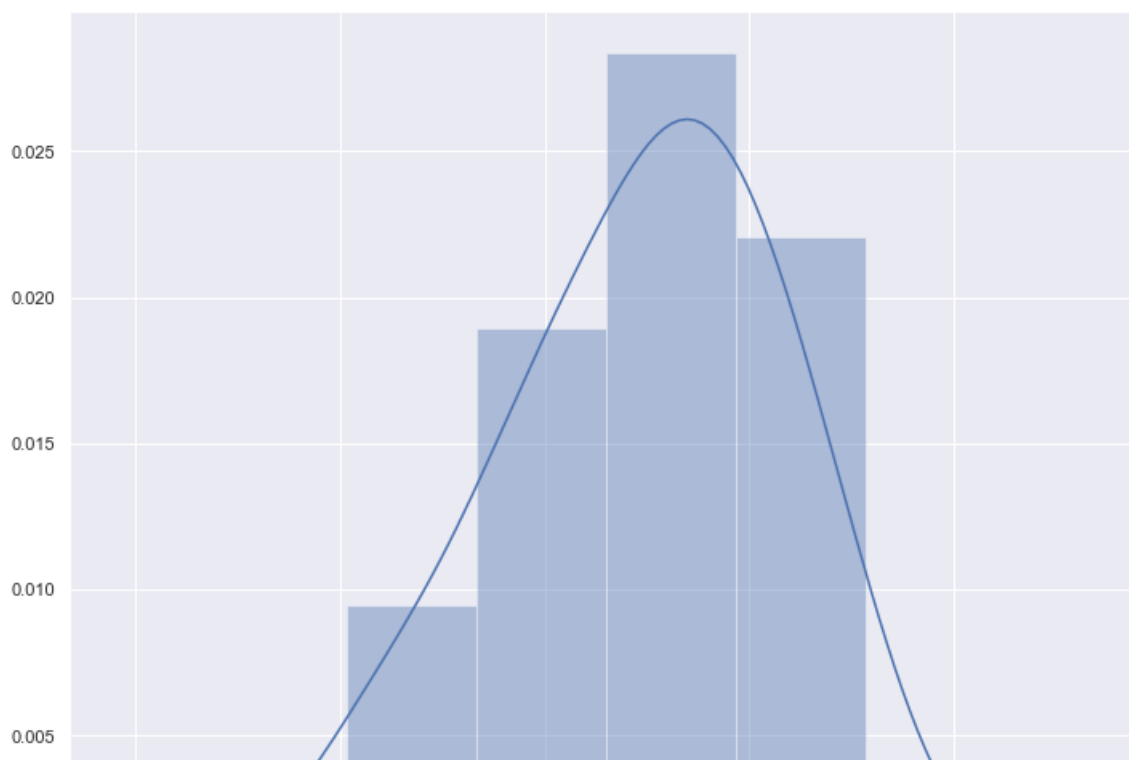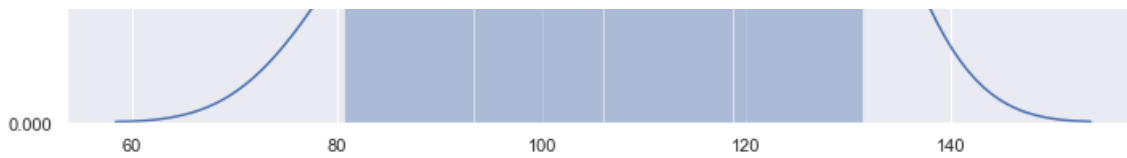Is the sample roughly normally distributed?

In [2]:

```
sns.set(color_codes=True)
sns.set(rc={'figure.figsize':(12,10)})
sns.distplot(sample)
```

Out[2]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2667e6ef080>
```

Step 1: Write your null and alternative hypothesis statements

As you are trying to monitor a change in the sales performance after the training, the null-hypothesis addresses the fact that there is no change and sales performance before and after the training is exactly the same.

H_0: *The null hypothesis is that there is no difference in sales, so:*

> $H_0: \mu = \$100.$

This is the one that you are testing. Our alternate hypothesis should address the expected change in the sales performance i.e. the sales performance has increased and the mean of sales post-training is greater than 100.

H_1: *The alternative hypothesis is that there is a change i.e. the mean sales increased.*

> $H_1: \mu > \$100.$

Step 2: Choose a Significance Level (Alpha)

The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. Look at the following graphs for a better understanding:
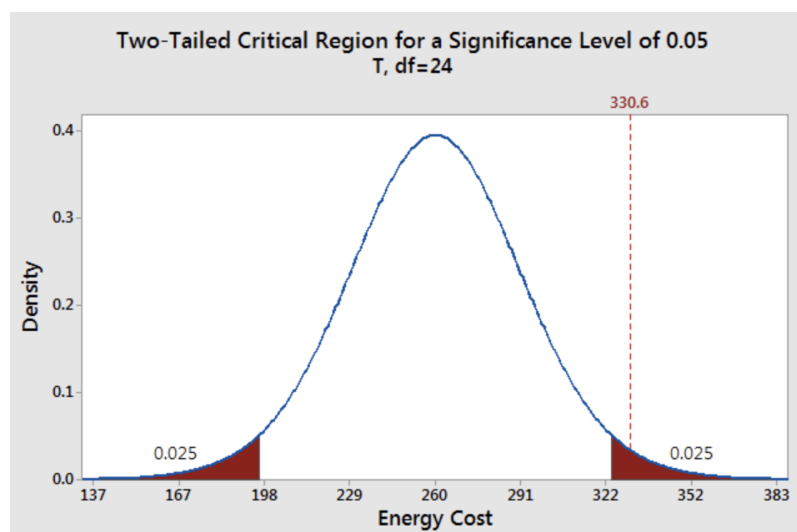

In the graph above, the two shaded areas are equidistant from the null hypothesis value and each area has a probability of 0.025, for a total of 0.05. In statistics, you call these shaded areas the critical regions for a two-tailed test. If the population mean is 260, you'd expect to obtain a sample mean that falls in the critical region 5% of the time. The critical region defines how far away our sample statistic must be from the null hypothesis value before you can say it is unusual enough to reject the null hypothesis.

Our sample mean (330.6) falls within the critical region, which indicates it is statistically significant at the 0.05 level.

You can also see if it is statistically significant using the other common significance level of 0.01.

0.005   0.005

0.0

137   167   198   229   260   291   322   352   383

Energy Cost
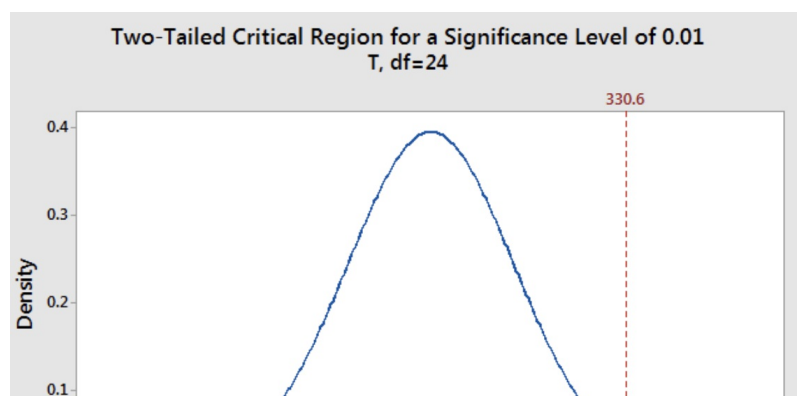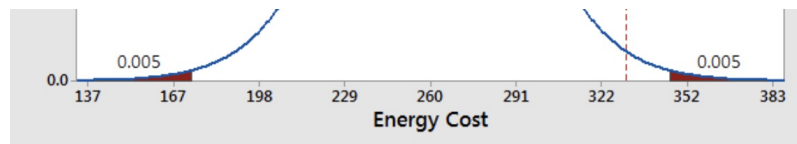
The two shaded areas each have a probability of 0.005, the two of which add up to a total probability of 0.01. This time the sample mean does not fall within the critical region, and you fail to reject the null hypothesis. This comparison shows why you need to choose your significance level before you begin your study. It protects you from choosing a significance level because it conveniently gives you significant results!

Using the graph, data scientists are able to determine that their results are statistically significant at the 0.05 level without using a p-value. However, when you use the numeric output produced by statistical software, you'll need to compare the p-value to your significance level to make this determination.

For Acme's experiment, you can assume an α of 0.05.

Step 3: Calculate the t-statistic

The sample looks like a nicely shaped normal distribution. After fulfilling the three requirements for a t-test mentioned above i.e. normality, independence, and randomness, we are ready to calculate our t statistic using the formula for one-sample t-test given as:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

> Using the formula given above, calculate the t-value in Python:

In [3]:

```
# Calculate Sigma
t = (x_bar -  mu)/(sigma/np.sqrt(n))
t
# 3.578139767278185
```

Out[3]:

```
3.578139767278185
```

The sample generated a t-statistic of around 3.58. Where in the t-distribution is this located? Let's try visualizing the calculated t-statistic on top of a PDF.

In [4]:

```
# generate points on the x axis between -5 and 5:
xs = np.linspace(-5, 5, 200)

# use stats.t.pdf to get values on the probability density function for the t-distribution
# the second argument is the degrees of freedom
ys = stats.t.pdf(xs, df, 0, 1)

# initialize a matplotlib "figure"
fig = plt.figure(figsize=(8,5))

# get the current "axis" out of the figure
ax = fig.gca()

# plot the lines using matplotlib's plot function:
ax.plot(xs, ys, linewidth=3, color='darkblue')

# plot a vertical line for our measured difference in rates t-statistic
```
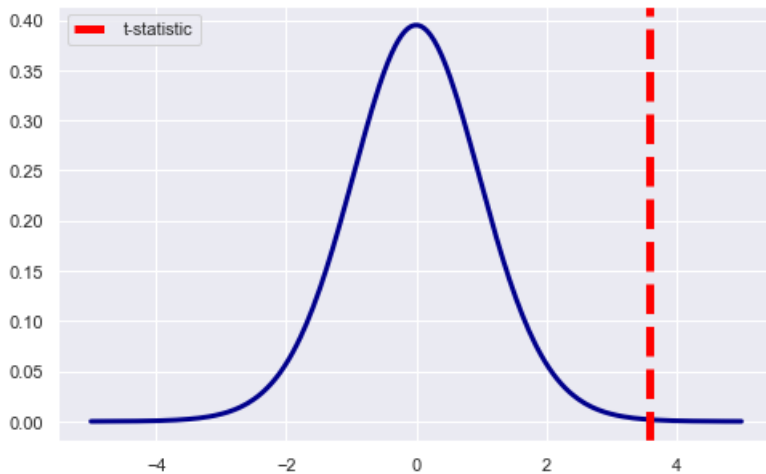
```
ax.axvline(t, color='red', linestyle='--', lw=5, label='t-statistic')
ax.legend()
plt.show()
```


Step 4: Calculate critical value (find rejection region)

Note that a positive t-value indicates that the sample mean is greater than the population mean and vice versa. This means that the sample's average sales performance post-training is greater than average population sales performance.

This sounds like good news, **BUT** is the increase high enough to reject the null hypothesis and accept that there is a significant increase in the mean of post-training sales performance, or is it just by chance. It's possible to calculate a critical t-value with a t-table and also by using Python `scipy.stats` module.

The critical value approach involves determining "likely" or "unlikely", by determining whether or not the observed test statistic is more extreme than would be expected if the null hypothesis were true. This involves comparing the observed test statistic to some cutoff value, called the **"critical value"**.

> If the test statistic is more extreme than the critical value, then the null hypothesis is rejected in favor of the alternative hypothesis. If the test statistic is not as extreme as the critical value, then the null hypothesis is not rejected.

You need two values to find this:

The **alpha level**: given as 5% in the question.

Degrees of freedom, which is the number of items in the sample (n) minus 1: 25 – 1 = 24.

t distribution critical values

df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
							Upper-tail probability p					
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725

	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

You use a one-tailed t-test towards the positive (right side of the t-distribution) to identify an increase in the sales performance.

Look up 24 degrees of freedom in the left column and a p-value of 0.05 (from 5% alpha level - 95% confidence level) in the top row. The intersection is `1.711`. This is our one-sample critical t-value.

For the Null hypothesis to be true, what this critical value means is that you would expect most values to fall under 1.711. If your calculated t-value (from Step 4) falls within this range, the null hypothesis is likely true and you would fail to reject the null hypothesis.

This value can also be calculated in Python using `scipy.stats` module using `ppf()` (Percent Point Function) as `scipy.stats.t.ppf(1-alpha, df)`.

Let's calculate the critical t using this formula and confirm our earlier findings.

In [5]:

```
# Calculate critical t-value
t_crit = np.round(stats.t.ppf(1 - 0.05, df=24),3)
t_crit
# 1.711
```

Out[5]:

1.711

As you can see, the critical value returned from the function (rounded off 2 to two decimal places) is the same as the one you should have found in the t-distribution table i.e. 1.711.

Using matplotlib, you can graph the rejection region as such. Any t-statistic that falls within the shaded region to the right will cause the hypothesis to be rejected.

In [6]:

```
# generate points on the x axis between -5 and 5:
xs = np.linspace(-5, 5, 200)

# use stats.t.pdf to get values on the probability density function for the t-distribution
# the second argument is the degrees of freedom
ys = stats.t.pdf(xs, df, 0, 1)

# initialize a matplotlib "figure"
fig = plt.figure(figsize=(8,5))

# get the current "axis" out of the figure
ax = fig.gca()

# plot the lines using matplotlib's plot function:
ax.plot(xs, ys, linewidth=3, color='darkblue')


ax.axvline(t_crit,color='green',linestyle='--',lw=4,label='critical t-value')
ax.legend()
ax.fill_betweenx(ys,xs,t_crit,where= xs > t_crit)
plt.show()
```
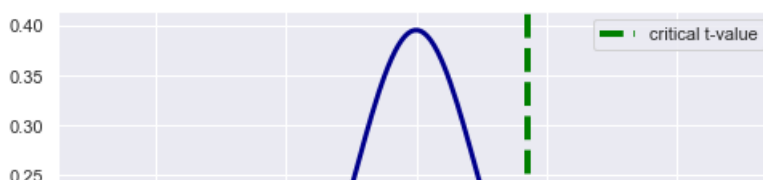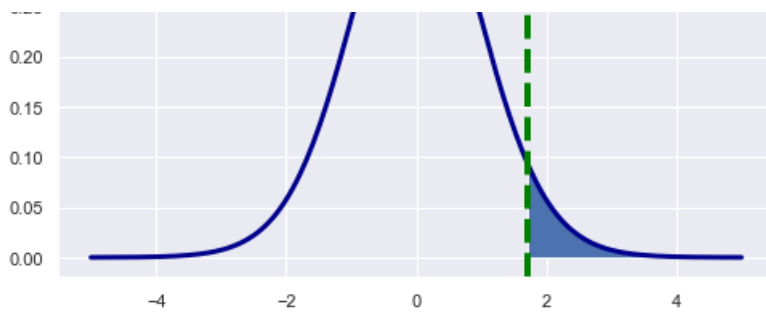
Step 5: Compare t-value with critical t-value to accept or reject the Null Hypothesis

Any t-value which is greater than 1.711 will fall into the shaded region in the above figure. t-values greater than 1.711 would reflect an "extreme" result and can be used to reject the null hypothesis.

Your calculated t-value, known as the t-statistic is 3.65, which is greater than 1.711 and hence the results can be called "statistically significant" and will allow researchers to reject the null hypothesis and with 95% confidence state that:

We are 95% sure that the mean sales performance post training is higher than the population mean prior to training.

NOTE: This calculation can also be performed using the `ttest_1samp` function in `SciPy.stats` indicated here:

> **scipy.stats.ttest_1samp(a, popmean, axis=0, nan_policy='propagate')**

Where a is the sample mean ($\bar{x}$) and popmean (μ) is the population mean. This function returns the t-value and p-value for the sample. Here, you are using a one-tailed t-test as you are looking for an increase in sales performance.

In [7]:

```python
results = stats.ttest_1samp(a= sample, popmean= mu)
print ("The t-value for sample is", round(results[0], 2), "and the p-value is", np.round((results[1]),
4))
#  Print results
# The t-value for sample is 3.58 and the p-value is 0.0015
```

The t-value for sample is 3.58 and the p-value is 0.0015

You can use our null and alternate hypotheses defined earlier to state the results from our findings.

In [8]:

```python
if (results[0]>t_crit) and (results[1]<0.05):
    print ("Null hypothesis rejected. Results are statistically significant with t-value =",
           round(results[0], 2), "and p-value =", np.round((results[1]), 4))
else:
    print ("Null hypothesis is Accepted")
```

Null hypothesis rejected. Results are statistically significant with t-value = 3.58 and p-value = 0.0015

It's also possible to visualize where the calculated t-statistic is compared to the critical t-value:

In [9]:

```python
# generate points on the x axis between -5 and 5:
xs = np.linspace(-5, 5, 200)

# use stats.t.pdf to get values on the probability density function for the t-distribution
# the second argument is the degrees of freedom
ys = stats.t.pdf(xs, df, 0, 1)

# initialize a matplotlib "figure"
fig = plt.figure(figsize=(8,5))

# get the current "axis" out of the figure
ax = fig.gca()

# plot the lines using matplotlib's plot function:
ax.plot(xs, ys, linewidth=3, color='darkblue')

# plot a vertical line for our measured difference in rates t-statistic
```
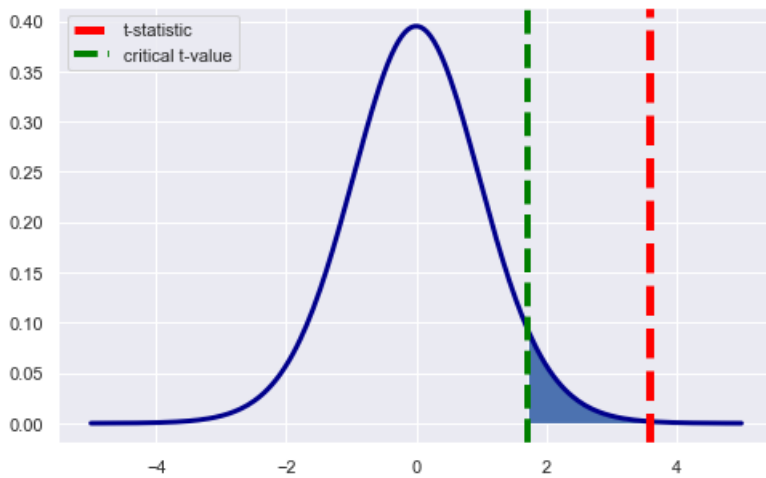
```
# plot a vertical line for our measured difference in rates t-statistic
ax.axvline(t, color='red', linestyle='--', lw=5,label='t-statistic')

ax.axvline(t_crit,color='green',linestyle='--',lw=4,label='critical t-value')
ax.fill_betweenx(ys,xs,t_crit,where= xs > t_crit)

ax.legend()
plt.show()
```


Summary

In this lesson, you saw a quick introduction to hypothesis testing using frequentists methods with t-values and p-values. You saw how a one sample t-test can be applied to contexts where the population mean is unknown and you have a limited amount of sample data. You looked at all the stages required for such hypothesis testing with a description of steps and also, how to perform these steps in Python.