z-statistics (z-score) is used when the data follows a normal distribution, population standard deviation is known and the sample size is above 30. Z-Score tells you how many standard deviations from the mean the result is.

t-statistics (t-score), also known as Student's T-Distribution, is used when the data follows a normal distribution, population standard deviation (σ) is NOT known, but the sample standard deviation (s) is known or can be calculated, and the sample size is below 30. T-Score tells you how many standard deviations from the mean the result is.

z-critical = stats.norm.ppf(1 - alpha)

t-critical = stats.t.ppf(alpha/numOfTails, ddof)

1) A sample of 20 items is selected randomly from a very large shipment. It is found to have a mean weight of 310 gm and standard deviation equal to 9 gm. Derive the 95% and 99% confidence intervals for population mean weight. ¶

In [1]:
```
from scipy import stats
import numpy as np
import math
import statistics as st
```

xbar = 310 n = 20 std dev = 9 Assuming weight of the items is normally distributed, we will have the C.I for mean determined by using t statistic. sample size less than 30 and also population standard deviation is not known so we use t-test.

In [2]:
```
n=20
std_dev = 9
```

In [3]:
```
std_error = std_dev/math.sqrt(n)
std_error
```

Out[3]:
```
2.0124611797498106
```

In [4]:
```
#scipy.stats.t.interval(confidence_level, degrees_freedom, sample_mean, sample_standard_error)

stats.t.interval(.95, n-1, 310, std_error) #95% confidence interval
```

Out[4]: (305.7878703422208, 314.2121296577792)

In [5]:
```
stats.t.interval(.99,n-1,310,std_error) #99% Confidence interval
```

Out[5]: (304.24248016671675, 315.75751983328325)

95% confidence interval : (305.7878703422208, 314.2121296577792) 99% Confidence interval : (304.24248016671675, 315.75751983328325)

2) 10 bars of a certain quality are tested for their diameters. The results are given below. Construct a 95 confidence interval of the mean diameters of the bars produced by the process.

diameter(in cm) : 1.02, 0.98, 0.97, 1.01, 0.94, 0.98, 1.00, 1.03, 0.92,1.02

sample size less than 30

In [6]:
```
x= np.array([1.02,0.98,0.97,1.01,0.94,0.98,1.00,1.03,0.92,1.02])
```

In [7]:
```
sample_mean = st.mean(x)
sample_mean
```

Out[7]: 0.987

In [8]:
```
sample_std_dev = st.stdev(x)
sample_std_dev
```

Out[8]: 0.036224607965059094

In [9]:
```
std_error = sample_std_dev/math.sqrt(len(x))
std_error
```

Out[9]: 0.01145522685162639

In [10]:
```
stats.t.interval(0.95,len(x)-1,sample_mean,std_error) #95% confidence interval
```

Out[10]:
```
(0.9610864765267704, 1.0129135234732296)
```

95% confidence interval : (0.9610864765267704, 1.0129135234732296)

3) Obtain a 90% confidence interval for the mean of population from which a sample of size 10 is drawn randomly. The population is known to be normally distributed and has a standard devation 15. The sample drawn has been found to have a mean equal to 42.

n=10, xbar = 42, std dev=15 Population standard deviation is known so use Z test

In [11]:
```
n=10
mean = 42
std_dev = 15
std_error = std_dev/math.sqrt(n)
```

In [12]:
```
z_critical=stats.norm.ppf(1 - .05)  #z alpha/2
z_critical                          #alpha= (1 - confidenceLevel)/numOfTails
```

Out[12]:
```
1.6448536269514722
```

In [13]:
```
lowerCI = mean - (z_critical * std_error)
upperCI = mean + (z_critical * std_error)
```

In [14]:
```
print(lowerCI, upperCI)
```

```
34.19777418186664 49.80222581813336
```

90% confidence interval is : 34.19777418186664 , 49.80222581813336

4) The quality-control manager at a Li-BATTERY factory needs to determine whether the mean life of a large shipment of Li-Battery is equal to the specified value of 375 hours. The process standard deviation is known to be 100 hours. A random sample of 64 batteries indicates a sample mean life of 350 hours. At the 0.05 level of significance is there evidence that the mean life is different from 375 hours? Computed the p-value is? at 95% confidence interval estimate of the population mean life of the battery is?

H0: Null Hypothesis, Mu0 = 375 H1: alt. hypothesis, Mi1 != 375 xbar = 350, std dev = 100, n = 64

In [15]:
```
n=64
Mu = 375
mean = 350
std_dev = 100
std_error = std_dev/math.sqrt(n)
```

In [16]:
```
z_critical=stats.norm.ppf(1 - .05/2)  #z alpha/2
z_critical                           #alpha= (1 - confidenceLevel)/numOfTails
```

Out[16]:
```
1.959963984540054
```

In [17]:
```
lowerCI = mean - (z_critical * std_error)
upperCI = mean + (z_critical * std_error)
```

In [18]:
```
print(lowerCI, upperCI)
```

```
325.5004501932493 374.4995498067507
```

At 95% Confidence interval estimate of the population mean life is 325.50 to 374.49

In [19]:
```
#z_score = (xbar - Mu)/(std_dev/sqrt(n))

z_score = (350 - 375)/(100/math.sqrt(64))
```

In [20]:
```
z_score
```

Out[20]:
```
-2.0
```

In [21]:
```
stats.norm.sf(abs(z_score))*2  #calculating p_value
                               #multiplied by 2, since 2 sided test
```

Out[21]:
```
0.04550026389635839
```

Calculated |z|=2.0 is greater than z critical value 1.96 o the null hypothesis is rejected. Also p value = 0.04 is less than alpha = .05 so we reject null hypothesis. We reject the null hypothesis that the population mean is 375 hours.

5) The mean cost of a hotel room in a city is said to be 168 dollars per night. A random sample of 25 hotels resulted in X-bar = 172.50 dollars and sample standard deviation s = 15.40. Calculate the t statistic. At the alpha = 0.05 level what should you do with null hypothesis.

In [22]:
```
x_bar = 172.5
Mu = 168
std_dev = 15.40
alpha = 0.05
n=25
```

In [23]:
```
# calculate t_critical for two tailed test
left_tailed_t =stats.t.ppf(alpha/2, n-1) #stats.t.ppf(alpha/numOfTails, ddof)
right_tailed_t =stats.t.ppf(1- alpha/2, n-1)
print(left_tailed_t, right_tailed_t)
```

```
-2.063898561628021 2.0638985616280205
```

Whenever you perform a two-tailed test, there will be two critical values. In this case, the T critical values are 2.0639 and -2.0769. Thus, if the test statistic is less than -2.0639 or greater than 2.0639, the results of the test are statistically significant(rejection of null hypothesis).

In [24]:
```
t_score =(x_bar - Mu)/(std_dev/math.sqrt(n))
t_score
```

Out[24]:
```
1.461038961038961
```

|t| = 1.46 is less than t critical value 2.06, so we cannot reject null hypothesis. Statistically insignficant.

In [25]:
```
p_val = stats.t.sf(abs(t_score), n-1)*2  # calculate p val
p_val                                    #multiplied by 2 since 2 tailed test
```

Out[25]:
```
0.1569743731561739
```

When your p-value is less than or equal to your significance level, you reject the null hypothesis. The data favors the alternative hypothesis. Results are statistically significant. When your p-value is greater than your significance level, you fail to reject the null hypothesis. Results are not significant. Here p value(0.15) is greater than alpha(0.05) so we cannot reject null hypothesis.