ASSIGNMENT 2

ESTIMATION AND CONFIDENCE INTERVALS

BASIC STATISTICS LEVEL 2

Scenario

A manufacturer of print-heads for personal computers is interested in estimating the mean durability of their print-heads in terms of the number of characters printed before failure. To assess this, the manufacturer conducts a study on a small sample of print-heads due to the destructive nature of the testing process.

Data

A total of 15 print-heads were randomly selected and tested until failure. The durability of each print-head (in millions of characters) was recorded as follows:

1.13, 1.55, 1.43, 0.92, 1.25, 1.36, 1.32, 0.85, 1.07, 1.48, 1.20, 1.33, 1.18, 1.22, 1.29

Assignment Tasks

a. Build 99% Confidence Interval Using Sample Standard Deviation

Assuming the sample is representative of the population, construct a 99% confidence interval for the mean number of characters printed before the print-head fails using the sample standard deviation. Explain the steps you take and the rationale behind using the t-distribution for this task.

b. Build 99% Confidence Interval Using Known Population Standard Deviation

If it were known that the population standard deviation is 0.2 million characters, construct a 99% confidence interval for the mean number of characters printed before failure.

```
import numpy as np
import pandas as pd
from scipy import stats

import warnings
warnings.filterwarnings("ignore")
```

So we imported the libraries and know we use store the given data points into a Data Frame and use the describe method to obtain the mean and standard deviation in one go

```
df.describe()
```

	0
count	15.000000
mean	1.238667
std	0.193164
min	0.850000
25%	1.155000
50%	1.250000
75%	1.345000
max	1.550000

so we get to know the mean and std of the data given here

so to find the confidence interval we have two methods to follow

- z test
- t test

we use z test when we know the population std deviation

we use t test when we dont know population std deviation

but, when the smaple size is large enough (30 samples are considered large enough) then we can use any one between z & t tests

for the z test the confidence interval theoritically is,

C.I = MEAN +/- Z(1-a).(std deviation/sqrt(n))

- here z(1-a) is the value of z scores at the level of confidence required and that can be calculated by
    - z= (x-mu)/sigma

for the t test the confidence inetrval theoritically is

C.I = MEAN +/- T(1-a,n-1).(std deviation/sqrt(n))

- here t(1-a,n-1) is the value of t scores at the level of confidence required and that can be calculated by
- t=(x-x'bar)/(std deviation/sqrt(n))

and to tell if the size of the smaple is large enough then we use n>= 10*(skewness)^2

a. Build 99% Confidence Interval Using Sample Standard Deviation Assuming the sample is representative of the population, construct a 99% confidence interval for the mean number of characters printed before the print-head fails using the sample standard deviation. Explain the steps you take and the rationale behind using the t-distribution for this task.

```
[ ] ci=stats.norm.interval(0.99,
                           loc=df.mean(),
                           scale=df.std()/np.sqrt(len(df)))
    print("mean at 99% confidence inetval is: ",np.round(ci,4))
```

```
mean at 99% confidence inetval is:  [[1.1102]
 [1.3671]]
```

the range of mean characters printed at a confidence of 99% is 1.11 mil to 1.36 mil using z test

```
[ ] a=stats.t.interval(0.99,df=14,loc=1.2387,scale=0.1932/(15**0.5))
    a

    #0.95 level of confidence, degrees of freedom (n-1), sample mean, sample deviation/sqrt(n)

    #here the sample is large enough so that we can rule out whether to use z method or t method, we can use either methods and it will be same or maybe a slight diffrence
```

```
(1.09020310114456, 1.387196898855439 7)
```

the range of mean characters printed at a confidence of 99% is 1.09 mil to 1.38 mil using t test

b. Build 99% Confidence Interval Using Known Population Standard Deviation If it were known that the population standard deviation is 0.2 million characters, construct a 99% confidence interval for the mean number of characters printed before failure.

```
ci_n=stats.norm.interval(0.99,
                         loc=df.mean(),
                         scale=0.2/np.sqrt(len(df)))
print("mean at 99% confidence inetval is: ",np.round(ci,4))
```

```
mean at 99% confidence inetval is:  [[1.1057]
 [1.3717]]
```

```
a_n=stats.t.interval(0.99,df=14,loc=1.2387,scale=0.2/(15**0.5))
a_n
```

```
(1.0849765022200415, 1.3924234977799583)
```

the range of mean characters printed at a confidence of 99% with 0.2 standard deviation is 1.08 mil to 1.39 mil