Assignment z-test

Date: 05/05/2020 Name: D.Saravanan

1. A sample of 400 male students is found to have a mean height of 171.38 cm. Can it be reasonably regarded as a sample from a large population with mean height 171.17 cm and standard deviation 3.30 cm.

Solution:

Given that $n = 400$, $\bar{x} = 171.38$ cm, $\mu = 171.17$ cm, $\sigma = 3.30$ cm

Sample proportion:

$$p = \frac{\bar{x}}{n} = \frac{171.38}{400} = 0.42845$$

Null hypothesis: $H_0 : \mu = 171.17$
Alternative hypothesis: $H_1 : \mu \neq 171.17$ (two tailed test)
z-statistic:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{171.38 - 171.17}{3.30/\sqrt{400}} = 1.27273$$

At 5% significance level the tabulated value for z_α is 1.96

Conclusion: $|z| < z_\alpha$, we accept the Null hypothesis. That is there is no significant difference between the sample mean and the population mean.

95% fiducial limits (Confidence Interval):
The confidence interval for the population mean is given by

$$\bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$171.38 \pm 1.96 \frac{3.30}{\sqrt{400}} = 171.38 \pm 0.3234$$

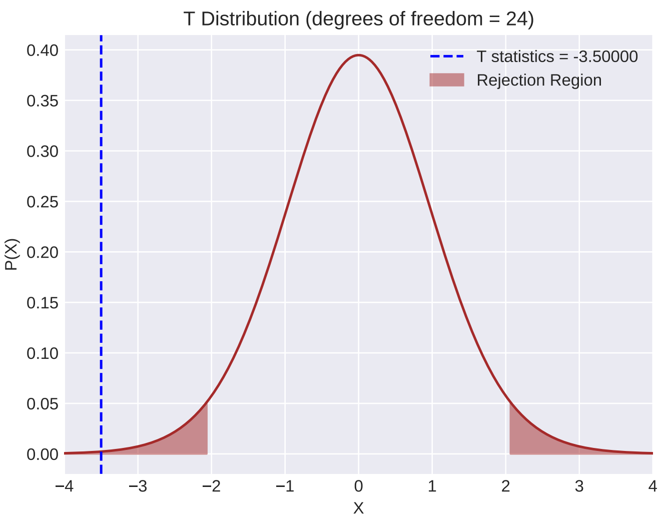
Program:

```
#!/usr/bin/env python3
import numpy as np
from scipy.stats import t
import matplotlib.pyplot as plt
plt.style.use('seaborn-darkgrid')

print("__________T-test__________")

p_mean = float(input("\nPopulation mean: "))
s_mean = float(input("Sample mean: "))
s_stdv = float(input("Sample stand_dev: "))
n = float(input("Sample length: "))

# degrees of freedom, level of significance, confidence level
dof = n - 1; los = 0.05; cnl = 1 - los

# calculation of T-statistics
tstatistics = (s_mean - p_mean)/(s_stdv/np.sqrt(n))
tstatistics = -tstatistics if tstatistics > 0 else tstatistics
print("\nT-statistics: {:.5f}".format(tstatistics))

# calculation of Critical values
tcritical_l = t.ppf(q = los/2, df = dof)
tcritical_u = -tcritical_l
print("\nCritical values are {:.5f}, {:.5f}".format(tcritical_l, tcritical_u))

# decision making - using T-statistics and Critical values
if tstatistics < tcritical_l or tstatistics > tcritical_u:
    print("Reject the Null hypothesis.")
else: print("Fail to reject the Null hypothesis.")
```

```
# calculation of p-value
pvalue = 2*t.cdf(tstatistics, df = dof)
print("\np-value: {:.5f}".format(pvalue))

# decision making - using p-value
if pvalue < los: print("Reject the Null hypothesis.")
else: print("Fail to reject the Null hypothesis.")

# standard error
std_err = s_stdv/np.sqrt(n)
print("\nStandard Error: {:.5f}".format(std_err))

# confidence interval
cnf_int = s_mean + std_err * np.array([tcritical_l, tcritical_u])
print("Confidence Interval: {}".format(cnf_int))

# plot script
x1=np.linspace(-10,tcritical_l,1000); y1=t.pdf(x1,df=dof)
x2=np.linspace(tcritical_u,10,1000);  y2=t.pdf(x2,df=dof)
x3=np.linspace(-4,4,1000); y3=t.pdf(x3,df=dof)
ax=plt.figure().add_subplot(111); ax.plot(x3,y3,color='brown')
ax.fill_between(x1,y1,0,alpha=0.5,color='brown',label='Rejection Region')
ax.fill_between(x2,y2,0,alpha=0.5,color='brown'); ax.set(xlabel='X',ylabel='P(X)')
ax.axvline(x=tstatistics,ls='--',c='b',label='T statistics = {:.5f}'.format(tstatistics))
ax.set(xlim=[-4,4],title="T Distribution (degrees of freedom = {:.0f})".format(dof))
plt.legend(); plt.show()
```

Output:

```
__________T-test__________

Population mean: 16
Sample mean: 16.35
Sample stand_dev: 0.5
Sample length: 25

T-statistics: -3.50000

Critical values are -2.06390, 2.06390
Reject the Null hypothesis.

p-value: 0.00184
Reject the Null hypothesis.

Standard Error: 0.10000
Confidence Interval: [16.14361014 16.55638986]
```



2. A sample of 900 items has mean 3.4 and standard deviation 2.61 Can the sample be regarded as drawn from a population with mean 3.25 at 1 percent level of significane.

Solution:

Null hypothesis: $H_0 : \mu = 3.25$
Alternative hypothesis: $H_1 : \mu \neq 3.25$ (two tailed test)

Program:

```
#!/usr/bin/env python3
import numpy as np
from scipy.stats import t, ttest_1samp
import matplotlib.pyplot as plt
plt.style.use('seaborn-darkgrid')

print("__________T-test (One Sample)__________")

p_mean = float(input("\nPopulation mean: "))
n = int(input("Sample length: "))
sample = [float(value) for value in input("Sample values: ").split()]
s_mean = np.mean(sample); s_stdv = np.std(sample, ddof=1)

# degrees of freedom, level of significance, confidence level
dof = n - 1; los = 0.05; cnl = 1 - los

# calculation of T-statistics and p-value
tstatistics, pvalue = ttest_1samp(sample, p_mean)
print("\nT statistics: {:.5f}".format(tstatistics))

# calculation of Critical values
tcritical_l = t.ppf(q = los/2, df = dof)
tcritical_u = -tcritical_l
print("\nCritical values are {:.5f}, {:.5f}".format(tcritical_l, tcritical_u))

# decision making - using T-statistics and Critical values
if tstatistics < tcritical_l or tstatistics > tcritical_u:
    print("Reject the Null hypothesis.")
else: print("Fail to reject the Null hypothesis.")

print("\np-value: {:.5f}".format(pvalue))

# decision making - using p-value
if pvalue < los: print("Reject the Null hypothesis.")
else: print("Fail to reject the Null hypothesis.")

# standard error
std_err = s_stdv/np.sqrt(n)
print("\nStandard Error: {:.5f}".format(std_err))

# confidence interval
cnf_int = s_mean + std_err * np.array([tcritical_l, tcritical_u])
print("Confidence Interval: {}".format(cnf_int))

# plot script
x1=np.linspace(-10,tcritical_l,1000); y1=t.pdf(x1,df=dof)
x2=np.linspace(tcritical_u,10,1000);  y2=t.pdf(x2,df=dof)
x3=np.linspace(-4,4,1000); y3=t.pdf(x3,df=dof)
ax=plt.figure().add_subplot(111); ax.plot(x3,y3,color='brown')
ax.fill_between(x1,y1,0,alpha=0.5,color='brown',label='Rejection Region')
ax.fill_between(x2,y2,0,alpha=0.5,color='brown'); ax.set(xlabel='X',ylabel='P(X)')
ax.axvline(x=tstatistics,ls='--',c='b',label='T statistics = {:.5f}'.format(tstatistics))
ax.set(xlim=[-4,4],title="T Distribution (degrees of freedom = {:.0f})".format(dof))
plt.legend(); plt.savefig('tscript2.pdf',dpi=72,bbox_inches='tight'); plt.show()
```

Output:

```
__________T-test (One Sample)__________

Population mean: 5
Sample length: 10
Sample values: 4.7 4.9 5.0 5.1 5.4 5.2 4.6 5.1 4.6 4.7

T statistics: -0.80472

Critical values are -2.26216, 2.26216
Fail to reject the Null hypothesis.

p-value: 0.44172
Fail to reject the Null hypothesis.

Standard Error: 0.08699
Confidence Interval: [4.73322266 5.12677734]
```