```python
"""
Created on Sat Apr  1 15:52:53 2017
Assignment 1
@author: Ho Wing Wong
"""

from scipy import stats
import scipy as sci
import numpy as np

def t_test (mean1, mean2, std1, std2, n1, n2, H0):
    t = (mean1-mean2 - H0)/((std1**2)/n1+(std2**2)/n2)**(1/2)
    return t

def degreeoffreedom (std1, std2,n1, n2):
    v1 = (std1**2)/n1
    v2 = (std2**2)/n2
    df = (v1+v2)**2/((v1**2)/(n1-1)+(v2**2)/(n2-1))
    return df

#11.2a
'''
u1 =  scenic route
u2 = the nonscenic route
H0: u1 - u2 = 10
Ha: u1 - u2 > 10
'''

#11.6a
'''
The two sample are independently choosen random samples
Sample size are both large
'''

#11.8
mean1 = 3.12
mean2 = 3.23
n1 = 184
n2 = 114
std1 = 0.4855
std2 = 0.524
t = t_test(mean1, mean2, std1, std2, n1, n2, 0)
df = degreeoffreedom(std1, std2, n1, n2)
p = stats.t.sf(t, df)
```

```python
'''
u1 = mean of GPA that are employed
u2 = mean of GPA that is not employed
H0: u1 - u2 = 0
Ha: u1 - u2 <0
t = -1.8
df = 225
p = 0.035739989393406724

Because p < 0.05, we reject H0.

'''

#11.9a
Dry = [302.1, 339.2, 288.8, 306.8, 305.2, 327.5]
Wet = [363.5, 377.7, 327.7, 331.9, 338.1, 394.6]
mean1 = sci.mean(Dry)
mean2 = sci.mean(Wet)
n1 = 6
n2 = 6
std1 = sci.std(Dry,ddof = 1)
std2 = sci.std(Wet, ddof = 1)
t = t_test(mean1, mean2, std1, std2, n1, n2, 0)
df = degreeoffreedom(std1, std2, n1, n2)
#Interval
a = (mean1-mean2 )+ (1.86*((std1**2)/n1+(std2**2)/n2)**(1/2))
b = (mean1-mean2 )- (1.86*((std1**2)/n1+(std2**2)/n2)**(1/2))
(a,b)
'''

mean1 = 311.59
mean2 = 355.58
df = 8.77
t (p(0.90), df = 8) =  1.86
90% interval : (-69, -19)

'''

#11.12a
Male = [1.3, 1.3, 0.9, 2.1, 0.7, 1.3, 3, 1.3, 0.6, 2.1]
Female = [-0.2, 0.5, 1.1, 0.7, 1.1, 1.2, 0.1, 0.9, 0.5, 0.5]
mean1 = sci.mean(Male)
mean2 = sci.mean(Female)
n1 = len(Male)
n2 = len(Female)
std1 = sci.std(Male, ddof = 1)
```

```python
std2 = sci.std(Female, ddof = 1)
t = t_test(mean1, mean2, std1, std2, n1, n2, 0)
df = degreeoffreedom(std1, std2, n1, n2)
p = stats.t.sf(t, df)

'''
mean1 = 1.45
mean2 = 0.64
std1 = 0.73
std2 = 0.45
H0: u1 - u2 = 0
Ha: u1 - u2 >0
t = 2.99
df = 14.96
p = 0.0046

Because p < 0.01, we reject H0.

'''

#11.32
Water = np.array([0.9, 0.92, 1, 1.1, 1.2, 1.25, 1.25, 1.3, \
                  1.35, 1.4, 1.4, 1.5, 1.65, 1.7, 1.75, 1.8, \
                  1.8, 1.85, 1.9, 1.95])
Syrup = np.array([0.92, 0.96, 0.95, 1.13, 1.22, 1.2, 1.26, \
                  1.3, 1.34, 1.41, 1.44, 1.52, 1.58, 1.7, 1.8, \
                  1.76, 1.84, 1.89, 1.88, 1.95])
Diff = Syrup - Water
Diffmean = Diff.mean()
Diffstd = Diff.std(ddof = 1)
n = 20
t = (Diffmean - 0) /(Diffstd/(n**(1/2)))
df = n-1
p = stats.t.sf(t, df)*2
'''
Diffmean = 0.00399
Diffstd = 0.0347
ud = u1 - u2 = Mean Velocity diference
H0: ud = 0
Ha: ud != 0
t = 0.52
df = 19
p = 0.612

Because P > 0.01, we cannot reject H0
'''
```

```python
#11.36
Method1 = np.array([27.5, 41.3, 3.5, 24.3, 27, 17.7, 12, 20.9])
Method2 = np.array([34.4, 38.6, 3.5, 21.9, 24.4, 21.4, 11.8, 24.1])
Diff = Method1 - Method2
Diffmean = Diff.mean()
Diffstd = Diff.std(ddof = 1)
n = 8
t = (Diffmean - 0)/(Diffstd/(n**(1/2)))
df = n-1
p = 1 - stats.t.sf(t/2, df)
'''
Diffmean = -0.737
Diffstd = 3.526
ud = u1 - u2 = Difference between the amount of radiation
of Method1 and Method2
H0: ud = 0
Ha: ud != 0
t = -0.59
p = 0.71

Because P is = 0.71, We cannot reject H0
'''

#11.38

#a
Credit1997 = np.array([61.4, 65.3, 65.1, 65.9, 42.3, 60.4, 42.9])
Credit2002 = np.array([52.8, 74.5, 72.4, 61.9, 62.7, 53.5, 62.2])
Diff = Credit1997 - Credit2002
Diffmean = Diff.mean()
Diffstd = Diff.std(ddof = 1)
n = 7
t = (Diffmean - 0)/ (Diffstd/(n**(1/2)))
df = n-1
p = 1 - stats.t.sf(t/2, df)

'''
Diffmean = -5.24
Diffstd = 12.05
ud = u1 - u2 = Difference mean percentage of exams earning credit
H0: ud = 0
Ha: ud !=0
t = -1.15
p = 0.293
```

```
Because P > 0.05, we cannot reject H0.
'''


#c
'''
Yes, it is reasonable because
1. the samples are paired
2. the n sample differences can be viewed as a random sample
from a population of differences
3. the number of sample differences is large (generally at least 30)
or the population distribution of differences is approximately normal,
the paired t confidence interval for is approximately normal.
'''


#11.44
p1 = 57/80
p2 = 50/60
n1 = 80
n2 = 60
pc = (n1*p1 + n2*p2)/(n1+n2)
z = (p1-p2)/ ((pc*(1-pc)/n1)+ (pc*(1-pc)/n2))**(1/2)
p = stats.t.sf(z, 1000000000)
'''
p1 = 0.7125
p2 = 0.833333
pc = 0.7642
ud = u1 - u2 = differences between registered by phone and through onlin
H0: ud = 0
Ha: ud <0
z = -1.66
p = 0.048

Because P < 0.05, we reject H0
'''


#11.46
p1 = 0.38
p2 = 0.27
n1 = 57
n2 = 50
pc = (n1*p1 + n2*p2)/(n1+n2)
z = (p1-p2)/ ((pc*(1-pc)/n1)+ (pc*(1-pc)/n2))**(1/2)
p = stats.t.sf(z, 1000000000)

'''
p1 = 0.38
```

```
p2 = 0.27
pc = 0.3285
ud = u1 - u2 = differences between stocking and standard treatment
H0: ud = 0
Ha: ud >0
z = 1.2
p = 0.113

Because P > 0.05, we canot reject H0
'''
```

```
#11.51a
p1 = 26/120
p2 = 222/419
n1 = 120
n2 = 419
pc = (n1*p1 + n2*p2)/(n1+n2)

#Interval
a = (p1-p2) + 1.96*((pc*(1-pc)/n1)+ (pc*(1-pc)/n2))**(1/2)
b = (p1-p2) - 1.96*((pc*(1-pc)/n1)+ (pc*(1-pc)/n2))**(1/2)
(b, a)

'''
p1 = 0.21
p2 = 0.52
pc = 0.46
ud = u1 - u2 = differences between stocking and standard treatment
H0: ud = 0
Ha: ud >0
Interval: (-0.41, -0.212)

'''
```